

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 122.665 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22
Sequence: 1 gattctgttcgcctgcctcctg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ses:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	296	10	S74899
2	22	100.0	363	11	BV089294
3	22	100.0	363	11	BV097602
4	22	100.0	463	10	DB3064
5	22	100.0	695	11	BV089295
6	22	100.0	695	11	BV097603
7	22	100.0	727	10	MUSMAMEB
8	22	100.0	746	4	AF215889
9	22	100.0	752	4	AF215889
10	22	100.0	757	4	BOVAMLGNII
11	22	100.0	798	10	BC059090
12	22	100.0	799	10	MUSMAMEB
13	22	100.0	820	4	BOVAMLGNII
14	22	100.0	5562	9	AB091787
15	22	100.0	6452	9	AB091783
16	22	100.0	6451	4	AB091789
17	22	100.0	8004	9	AB091786
18	22	100.0	9384	10	AF294397
19	22	100.0	95826	10	AL805974

20	20.4	92.7	65	6	CO531502
21	20.4	92.7	399	10	RU07054
22	20.4	92.7	457	10	RU060564
23	20.4	92.7	476	6	AR452534
24	20.4	92.7	500	6	AR077245
25	20.4	92.7	500	6	AR120190
26	20.4	92.7	500	6	BD107909
27	20.4	92.7	649	9	BC074951
28	20.4	92.7	659	9	BC069118
29	20.4	92.7	753	10	RU060562
30	20.4	92.7	765	4	AB032194
31	20.4	92.7	778	4	SSU43405
32	20.4	92.7	780	10	U01245
33	20.4	92.7	789	4	AB032193
34	20.4	92.7	793	6	AR077243
35	20.4	92.7	793	6	AR120188
36	20.4	92.7	793	6	CQ731756
37	20.4	92.7	793	6	BD107907
38	20.4	92.7	793	9	HUMANMGXA
39	20.4	92.7	812	10	RU067130
40	20.4	92.7	825	10	RU051195
41	20.4	92.7	835	9	AF436849
42	20.4	92.7	885	10	CP0012200
43	20.4	92.7	5684	9	AB091785
44	20.4	92.7	5712	4	AB091793
45	20.4	92.7	6465	9	AB091781

ALIGNMENTS

RESULT 1
LOCUS S74899 296 bp mRNA ROD 09-MAY-1995
DEFINITION amelogenin {M194 transcript, alternatively spliced} [mice, enamel organ epithelia, mRNA Partial, 296 nt].
ACCESSION S74899
VERSION S74899.1 GI:802018
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE 1 (bases 1 to 296)
AUTHORS Simmer,J.P., Hu,C.C., Lau,B.C., Sarte,P., Slavkin,H.C. and Fincham,A.G.
TITLE Alternative splicing of the mouse amelogenin primary RNA transcript
JOURNAL Calcif. Tissue Int. 55 (4), 302-310 (1994)
MEDLINE 95120737
PubMed 7820782
REMARK Genbank staff at the National Library of Medicine created this entry [NCBI glibseq 159650] from the original journal article.
FEATURES
source
1..296
/organism="Mus sp."
/mol_type="mRNA"
/db_xref="taxon:10095"
1..296
/gene="amelogenin"
69..296
/gene="amelogenin"
/codon_start=1
/product="amelogenin"
/protein_id="AAB33093.1"
/db_xref="GI:802019"
/translation="MGTWILFACLLGAAPAMPUPPHGSPGYINLSYKSHSQAINTD
RTALVLTPIKWTQSMIRQPPSYGVRPMGWL"

ORIGIN

Query Match 100.0% ; Score 22; DB 10; Length 296;
Best Local Similarity 100.0% ; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GATTGTGTTGCTGCTCCTG	22	
Db	80	GATTGTGTTGCTGCTCCTG	101	
RESULT 2				
BV089294/c			363 bp	DNA
LOCUS				linear
DEFINITION				STS 15-OCT-2003
ACCESSION	BV089294			
VERSION	BP089294			
KEYWORDS				
SOURCE	BV089294.1	GI:37666773		
ORGANISM	STS.			
	Mus musculus (house mouse)			
REFERENCE				
AUTHORS	Mus musculus			
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
COMMENT	1 (bases 1 to 363)			
	Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,			
	McPherson,J.D., Foerzler,D. and Peltz,G.			
	Mus musculus SNPs			
	Unpublished (2003)			
FEATURES				
Source	Contact: Jonathan Usuka			
	Roche Palo Alto Genetics and Genomics Department			
	Roche Palo Alto			
	3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA			
	Tel: 6508555807			
	Email: Jonathan.Usuka@roche.com			
	Primer A: No primer submitted			
	Primer B: No primer submitted.			
	Location/Qualifiers			
	1..363			
	/organism="Mus musculus"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:10090"			
	/map="x-73268-72907-AL0805974.7.1.95826"			
	/clone_lib="Roche Palo Alto"			
	/note="SNPs developed from assay sequences derived from 15			
	different strains-of mice (as of October 1, 2003). Those			
	strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D2-H2/OSu,			
	BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J~CAST/Ei, DBA/2J,			
	MRL/MpJ, NZB/B1mJ, NZW/Lac, SPRET/Ei."			
	<1..>363			
STS				
ORIGIN				
Query Match	100.0%;	Score 22;	DB 11;	Length 363;
Best Local Similarity	100.0%;	Pred. No. 1.9;		
Matches	22;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	
QY	1	GATTGTGTTGCTGCTCCTG	22	
Db	134	GATTGTGTTGCTGCTCCTG	113	
RESULT 3				
LOCUS	BV097602		363 bp	DNA
DEFINITION	RPAMSE00010112 Roche Palo Alto Mus musculus STS genomic, sequence			
ACCESSION	BV097602			
VERSION	BV097602			
KEYWORDS				
SOURCE	BV097602.1	GI:40810678		
ORGANISM	STS.			
	Mus musculus (house mouse)			
REFERENCE				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 363)			
	Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,			
	McPherson,J.D., Foerzler,D. and Peltz,G.			
	Mus musculus SNPs			
	Unpublished (2003)			
TITLE				
JOURNAL				
COMMENT				

Contact: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: jonathan.usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.

FEATURES
Source

1. 363
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73268-72907-AL005974.7.1.95826"
/clone_lib="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/OSu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MpJ, NZB/BlmJ, NZW/Lac, SPRET/Ei."

Query Match 100.0%; Score 22; DB 11; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
STS

1 GATTTGTTTGCCGTCGCTCGG 22
|||||
134 GATTTGTTTGCCGTCGCTCGG 113

Db

RESULT 4
LOCUS D83063S2 463 bp DNA linear ROD 26-NOV-2003
DEFINITION Mus musculus DNA for amelogenin, exon 2.
ACCESSION D83064
VERSION D83064.1 GI:2687869
KEYWORDS 2 of 5
SEGMENT Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Oida, S., Miyazaki, H., Iimura, T., Suzuki, M., Sasaki, S. and Shimokawa, H.
Molecular structure of the mouse amelogenin genomic DNA
DNA Seq. 6 (5), 307-310 (1996)
97142134
8988368
2 (bases 1 to 463)
Oida, S.
Direct Submission
Submitted (14-JAN-1996) Shinichi Oida, Tokyo Medical and Dental University, Fac. Dentistry, Dept. Biochemistry, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (E-mail:s.oida.obchdent.tmd.ac.jp, tel:03-5803-5448, fax:03-5803-0187)
On Dec 15, 1997 this sequence version replaced gi:2662357.
Location/Qualifiers
1. 463
/organism="Mus musculus"
/mol_type="genomic DNA"
/sub_species="Balb-c"
/db_xref="taxon:10090"
/chromosome="X"
/map="Xp1-F4"
/clone="MAMEG-2"
/cell_type="Leukocyte"
/clone_lib="EMBL-3 SP6-T7"
/dev_stage="adult"
join(D83063.1:203..258,222..233)
c1..221

COMMENT
FEATURES
Source

5'UTR
Initiation

exon /number=1
222..287
/product="amelogenin"

Intron /number=2
288..463
/number=2

ORIGIN

Query Match 100.0%; Score 22; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTCCTG 22
|||||
Db 245 GATTTGTTTGCTGCTCCTG 266

RESULT 5
BV089295/c 695 bp DNA linear STS 15-OCT-2003
LOCUS RPAMMSE0001191 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION tagged site.

ACCESSION BV089295 GI:3766774
VERSION BV089295
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 695)
Uauka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
Mus musculus SNPs
TITLE Unpublished (2003)
JOURNAL

COMMENT
Contact: Jonathan Uauka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Uauka@roche.com
Primer A: No primer submitted
Primer B: No primer submitted.
Location/Qualifiers

FEATURES
source 1..695
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73613-72919-AL805974.7.1.95826"

/clone_11b="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains-of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D-H2/OSuJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MLR/MpJ, NZB/BinJ, NZM/Lac, SPRET/Ei.."
<1..>695

ORIGIN

Query Match 100.0%; Score 22; DB 11; Length 695;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTCCTG 22
|||||
Db 479 GATTTGTTTGCTGCTCCTG 458

RESULT 6
BV097603/c 695 bp DNA linear STS 14-JAN-2004
LOCUS RPAMMSE00010113 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV097603

VERSION BV097603.1 GI:40810679
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 695)
Uauka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
Mus musculus SNPs
TITLE Unpublished (2003)
JOURNAL

ORIGIN

COMMENT
Contact: Jonathan Uauka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Uauka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.
Location/Qualifiers

FEATURES

source 1..695
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73613-72919-AL805974.7.1.95826"
/clone_11b="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D-H2/OSuJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MLR/MpJ, NZB/BinJ, NZM/Lac, SPRET/Ei.."
<1..>695

ORIGIN

Query Match 100.0%; Score 22; DB 11; Length 695;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTCCTG 22
|||||
Db 479 GATTTGTTTGCTGCTCCTG 458

RESULT 7
MUSMAMB 727 bp mRNA linear ROD 04-FEB-2003
LOCUS Mus musculus MAME mRNA for amelogenin, complete cds.
DEFINITION D31769
VERSION D31769.1 GI:1321652

KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Snead,M.L., Lau,E.C., Zeichner-David,M., Pincham,A.G., Woo,S.L. and
Slavkin,H.C.

TITLE DNA sequence for cloned cDNA for murine amelogenin reveal the amino
acid sequence for enamel-specific protein
JOURNAL Biochem. Biophys. Res. Commun. 129 (3), 812-818 (1985)
MEDLINE 85251692
PUBMED 4015654

REFERENCE

AUTHORS Lau,E.C., Simmer,J.P., Bringas,P. Jr., Heu,D.D., Hu,C.C.,
Zeichner-David,M., Thiemann,F., Snead,M.L., Slavkin,H.C. and
Pincham,A.G.

TITLE Alternative splicing of the mouse amelogenin primary RNA transcript
contributes to amelogenin heterogeneity
JOURNAL Biochem. Biophys. Res. Commun. 188 (3), 1253-1260 (1992)
MEDLINE 93075222
PUBMED 1445358
REFERENCE 3 (bases 1 to 727)

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Butheraia; Cecartiodactyla; Ruminantia; Pecora; Bovidae;
TITLE	Caprinae; Capra
JOURNAL	1 (bases 1 to 752)
REFERENCE	Chen,C.-M., Cheng,W.T.K., Chang,C.-C., Weng,T.-L., Jiang,Y.-N.
AUTHORS	Hsu,J.-T.
TITLE	Cloning, cDNA sequence, and alternative splicing of goat amelogenin
JOURNAL	mRNAs
FEATURES	Unpublished
SOURCE	2 (bases 1 to 752)
	Chen,C.-M., Cheng,W.T.K., Chang,C.-C., Weng,T.-L., Jiang,Y.-N. and
	Hsu,J.-T.
	Direct Submision
	Submitted (13-DEC-1999) Department of Zoology, Life Science
	College, No. 250, Kuo Kuang Road, Tainchung 402, Taiwan
	Location/Qualifiers
	1..752
	/organism="Capra hircus"
	/mol_type="mrna"
	/db_xref="taxon:9925"
	/chromosome="X"
	/clone="AML-X-10"
	/cell_type="ameloblast"
	/note="bred: Alpine"
	<1..>752
gene	/gene="AML-X"
5' UTR	<1..34
	/gene="AML-X"
CDS	35..658
	/gene="AML-X"
	/codon_start=1
	/product="amelogenin precursor"
	/protein_id="AAG4396.1"
	/db_xref="GI:12004312"
sig_peptide	/translation="MGTWILFACLLGAASMPRLPRHGPYINFSVEPTLKMYOSI MRIRPSYGYSVPMGMLHOIIIVSSQTPONALDPHHIIPVPAQQPVPOOPMMW PVPQHSMTPRQHQHPNLPLPAQRPQRQSIOPHNQLOSLQMQLPQLPQPQPP PVHTQPLPFPQPLPIFLPMDHLPEMLFDVLIEDPDATDKTKREVD"
	35..82
	/gene="AML-X"
	/note="secretory signal peptide"
3' UTR	659..752
	/gene="AML-X"
	/note="highly divergent from AML-Y 3' UTR"
ORIGIN	
Query Match	100.0%; Score 22; DB 4; Length 752;
Best Local Similarity	100.0%; Pred. No. 2;
Matches 22; Conservative 0; Mismatches	0; Indels 0; Gaps 0;
Dy	1 GATTTCCTGGCCCTGCCTCG 22
Db	46 GATTTCCTGGCCCTGCCTCG 67
RESULT 10	
LOCUS BOVMAMGNII	BOVMAMGNII 757 bp mRNA linear NAM 26-APR-1993
DEFINITION	Bovine (Class II) amelogenin mRNA, complete cds.
ACCSSION	M63500 J05307
VERSION	M63500.1 GI:162661
KYCCE	amelogenin.
SOURCE	Bos taurus (cow)
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Butheraia; Cecartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos. 1 (bases 1 to 757) Gibson,C., Gelub,E., Herold,R., Risser,M., Ding,W., Shimokawa,H., Young,M., Temblin,J. and Rosenblom,J. Structure and expression of the bovine amelogenin gene: Biochemistry 30 (4), 1075-1079 (1991) 9113686

FEATURES	COMMENT	ORIGINAL SOURCE TEXT: BOVINE FETAL TEETH AMELOBLAST (CLASS II), CDNA TO mRNA.
FEATURES	Location/Qualifiers	
SOURCE	1. .757	/organism="Bos taurus"
		/mol_type="mRNA"
		/db_xref="taxon:9913"
		/cell_type="ameloblast"
		/tissue_type="tooth"
		/dev_stage="fetus"
gene	1. .757	/gene="amelogenin"
	41. .619	/gene="amelogenin"
		/note="class II"
CDS		/codon_start=1
		/product="amelogenin"
		/protein_id="AAJ0373.1"
		/db_xref="GI:162662"
ORIGIN		/translation="MGWTLIPACILGAAPSPRLPRHGHGTYINFSEVUTPLKRYQS MIRHPYSVGYEPWGMGHLHQIIPVVSQTPQNALQPHHHPVPAQSPVPOQPMW PVPQHSMTPTQHQPNLPLPAQOSFPQPPQIPQPHQPPQVPHIQRLPPQPLP PIFPMQPLPVLPLDLPLEAMPATDVKTRREVD"
Query Match	100.0%;	Score 22; DB 4; Length 757;
Best Local Similarity	100.0%;	Pred. No. 2;
Matches 22; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1 GATTGTTGGCCGCGCTCGT 22	
Db	52 GATTGTTGGCCGCTCGT 73	
RESULT 11		
LOCUS	BC059090	798 bp mRNA linear ROD 20-OCT-2003
DEFINITION	Mus musculus cDNA clone MGC:70128 IMAGE:30286944, complete cds.	
ACCESSION	BC059090	
VERSION	BC059090.1	GI:3747841
KEYWORDS	MGC.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
AUTHORS	1 (bases 1 to 798)	
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Helix, F., Ditschenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheerz, T.E., Brownstein, M.J., Uedin, T.B., Tsuchiyuki, S., Canninci, P., Plange, C., Kaha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Boeck, S.A., McEwen, P.J., McEwan, K.C., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huylk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Mdan, A.C., Rodrigues, S., Sanchez, A., Whiting, M., Mdan, A., Young, A.C., Shvachenko, Y., Bouford, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, J., Skalska, U., Smalls, D.E., Suterch, A., Schein, J.E., Jones, S.J. and Marry, M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903 (2002)
MEDLINE	22388257	
PUBMED	12477932	
REFERENCE	2 (bases 1 to 798)	
AUTHORS	Strausberg, R.	
TITLE	Direct Submission	

JOURNAL

JOURNAL Submitted (01-OCT-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

REMARK
COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://Image.lnl.gov>
Series: IRAX Plate: 131 Row: d Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer Frequency ORF analysis, Genomescan gene prediction.

FEATURES

```
source
1. . 798
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon.10090"
/cclone="MGC:70128 IMAGE:30286944"
/tissue_type="Jaw and Limb, mouse, day 18.5 to new born,
pool of mature fore and hind limb, maxilla and mandible
containing endochondral and membranous bone, formed
joints, tendon, ligaments, dermis, epidermis, muscle and
teeth with newly forming dentin and enamel"
/cclone_1ib="NIN_MGC_136"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6.1"
39. . 698
DSDS
```

misc_feature

Amelogenin. Amelogenins play a role in biomineralization. They seem to regulate the formation of crystallites during the secretory stage of tooth enamel development, thought to play a major role in the structural organization and mineralization of developing enamel. They are found in the extracellular matrix. Mutations in X-chromosomal amelogenin can cause Amelogenesis imperfecta" db_xref="CDD:ptamo2948"

ORIGIN

Query Match	100.0%	Score 22	DB 10	Length 798
Best Local Similarity	100.0%	Pred. No. 2		
Matches	22	Conservative	0	Mismatches 0; Indels 0; Gaps 0;
Qy	1	GATTTTGTTGGCTGCCTGCTG	22	
Db	50	GATTTTGTTGGCTGCCTGCTG	71	

RESULT 12

MUSMAMEA	MUSMAMEA	799 bp	mRNA	linear	ROD 04-FEB-2003
LOCUS					
DEFINITION	Mus musculus	MAME	mRNA	for amelogenin, complete cds.	
ACCESSION	D31768				
VERSION	D31768.1	GI:1321650			
KEYWORDS					
SOURCE	Mus musculus				
ORGANISM	Mus musculus	(house mouse)			

REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED	REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED
1	Snead, M.L., Lau, E.C., Zeichner-David, M., Fincham, A.G., Woo, S.L. and Slavkin, H.C.				
	DNA sequence for cloned cDNA for murine amelogenin reveal the amino acid sequence for enamel-specific protein				
		85251692			
		4015654			
	Lau, E.C., Simmer, J.P., Bringas, P. Jr., Hsu, D.D., Hu, C.C., Zeichner-David, M., Thiemann, F., Snead, M.L., Slavkin, H.C. and Fincham, A.G.				
	Alternative splicing of the mouse amelogenin primary RNA transcript contributes to amelogenin heterogeneity				
		93075222			
		1445358			
	3 (bases 1 to 799)				
	Oida, S., Iimura, T., Arai, N., Takeda, K., Marnoka, Y., Terashima, T., Shimokawa, H. and Sasaki, S.				
	Unpublished				
	4 (bases 1 to 799)				
	Oida, S.				
	Direct Submission				
	Submitted (03-JUN-1994) Shinichiro Oida, Faculty of Dent., Tokyo Medical and Dental University, Department of Biochemistry; 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (Tel:81-3-3813-6111 (ex. 5126), Fax:81-3-5684-8047)				
	On May 16, 1996 this sequence version replaced gi:577326.				
	Location/Qualifiers				

exon

exon

CDS

```
sig_peptide      60..110
                 /gene="NAME"
mat_peptide      111..647
                 /gene="NAME"
```

	Query Match	Score 22;	DB 10;	Length 799;
	Best Local Similarity	100.0%;	Pred. No. 2;	
	Matches 22;	Conservative 0;	Mismatches 0;	Gaps 0;
QY	1 GATTTCGTTTGGCCCTCCTCG 22			
db	71 GATTTTGTTTGCCTGCCCTCG 92			

RESULT	13
LOCUS	BOVAMLGN1
DEFINITION	BOYAMT.GNT 820 bp mRNA linear MAM 26-APR-1993
ACCESSION	M63499 J05307
VERSION	M63499.1 GI:162659
KEYWORDS	amelogenin.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
REFERENCE	Boydinae, Bos. 1 (bases 1 to 820) Gibson,C., Golub,E., Herold,R., Risser,M., Ding,W., Shimokawa,H., Young,M., Termini,J. and Rosenblom,J. Structure and expression of the bovine amelogenin gene <i>Biochemistry</i> 30 (4), 1075-1079 (1991)
TITLE	JOURNAL
PUBMED	MELINE
COMMENT	Original source text: Bovine fetal teeth ameloblast (class I), cDNA to rRNA.

```

FEATURES
source
    Location/Qualifiers
        1..820
            /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            /cell_type="ameloblast"
            /tissue_type="tooth"
            /dev_stage="fetus"
            1..820
            /gene="amelogenin"
            41..682
            /gene="amelogenin"
            /note="class I"
            /codon_start=1
            /product="amelogenin"
            /protein_id="AA03072.1"
            /db_xref="GI:162660"
            /translation="MGTWILFACLLGAASFMPDLPPHPCHPGVINFSYEVLTPLKMYQSS
MIRHDPYSGYEGPMBGMLHNOILIPVSGOTPNALQDHNHILPMVPAQAPVPOOPMM
PVPGQSHMTPTQHNQNPILPLPAQAPFGQASIQPGNRLQDQNRQLQDQNPQNPQLQPLC
PLQAPQNPRIQLPAPQRLPPIFMQAPLPMPLPDLPLAAMPATDKTKREED"

```

RESULT 14	
AB091787	
LOCUS	AB091787
DEFINITION	Olelemur garmetiil AMELX gene for amelogenin, partial cds.
ACCESSION	AB091787
VERSION	AB091787.1 GI:29126027

SOURCE ORGANISM *Ocolemur garnettii* (small-eared galago)
Ocolemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Ocolemur.

REFERENCE AUTHORS TITLE
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

JOURNAL MEDLINE
22609859
12672862
2 (bases 1 to 5562)

REFERENCE AUTHORS TITLE
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokenai), Department of Biosystems Science, Shonan kokuaiimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

```
FEATURES
    source
        Location/Qualifiers
            1..5562
                /organism="Olelemur garnettii"
                /mol_type="genomic DNA"
                /db_xref="taxon:30611"
                /sex="male"
                267..5562
                    /gene="AMELX"
                    join(267..321,1401..1412)
                    /gene="AMELX"
                    join(1413..1466,3385..3432,4699..4740,4832..4876,
                        5125..>5562)
                    /gene="AMELX"
                    /codon_start=1
                    /product="amelogenin"
                    /protein_id="BAC6107.1"
                    /db_xref="GI:29126028"
                    /translation="MGWILLFASLLGAAFAFMRPLRRHRGHGVINFSYEKSAQNIDIT
                        KTLVLTPKYKYSMLRPRPYRSYGVEMSGMILNLTPTLSQGNRPNTTLOPHNIITVU
                        VMQAORVPDQRPMHVRFGDSMTRTIONHNLRRPRAODRPORVQDVQDRPDQRMQROE
                        MWMQRIQDPQRVNRMQRLRPPRLRFGLFMQRLVLRPLRLDLPLEAMDAT"
```

Query Match	100.0%;	Score 22;	DB 9;	Length 5562;
Best Local Similarity	100.0%;	Pred. No. 2.3;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1 GATTTTGTTCCTGCGCCTCCCTG 22 			
Dd	1424 GATTTCCTTGCGCCTCCCTG 1445 			
RESULT 15				
AB091783				
LOCUS	AB091783	6442 bp	DNA	linear PRI 02-MAY-2003
DEFINITION	Saimiri sciureus AMELX gene for amelogenin, partial cds.			
ACCESSION	AB091783			
VERSION	AB091783.1 GI:29126019			

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 29.1639 Seconds
(without alignment)
3957.220 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22
Sequence: 1 gattctgttcgtcgtcctccctg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23sep04:*
1: geneseqn1980a:*
2: geneseqn1990a:*
3: geneseqn2000a:*
4: geneseqn2001a:*
5: geneseqn2001b:*
6: geneseqn2002a:*
7: geneseqn2002b:*
8: geneseqn2003a:*
9: geneseqn2003b:*
10: geneseqn2003c:*
11: geneseqn2003d:*
12: geneseqn2004a:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	722	6	AAL41111
2	22	100.0	722	6	AAL41110
3	20.4	92.7	65	6	ABN28389
4	20.4	92.7	270	12	ADM80835
5	20.4	92.7	318	12	ADM80836
6	20.4	92.7	476	3	AAZ50832
7	20.4	92.7	549	12	ADM80826
8	20.4	92.7	556	10	ADB59026
9	20.4	92.7	556	10	ADB53782
10	20.4	92.7	623	12	ADM80827
11	20.4	92.7	750	2	AAZ07020
12	20.4	92.7	793	2	AAZ07018
13	20.4	92.7	259202	12	ADQ18492
14	20	90.9	20	2	AAI18413
15	19.4	88.2	2190	8	ABX16470
16	19.4	88.2	2190	8	ABX16470
17	19.4	88.2	2190	11	ADN39263
18	19.4	88.2	2685	6	AB167223
19	19.4	88.2	2685	6	AB167223
20	19.4	88.2	4055	10	ADP81652
21	19.4	88.2	4181	12	ADL12637

C	22	19	86.4	19	2	AAI18412	Aa18412 PCR prime
C	23	18.8	85.5	802	2	AAZ07019	Aa207019 Amelogeni
C	24	18.8	85.5	852	12	ADQ22977	Adq22977 Human sof
C	25	18.4	83.6	1287	4	ABK41746	Abk41746 cDNA enco
C	26	18.4	83.6	1287	9	ADB59413	Adb59413 Connectiv
C	27	18.4	83.6	1295	4	ABK42097	Abk42097 cDNA enco
C	28	18.4	83.6	1295	3	ADB59764	Adb59764 Connectiv
C	29	17.8	80.9	173	3	AAI31387	Aa31387 Human sec
C	30	17.4	79.1	80	12	ADM95374	Adm95374 Rat anti
C	31	17.4	79.1	116	2	AAZ00965	Aa200965 Pgi codin
C	32	17.4	79.1	3923	10	ABT42229	Abt42229 Toxicity
C	33	17.4	79.1	5324	2	AAZ00879	Aa200879 Pgi splic
C	34	17.4	79.1	56516	2	AAZ00870	Aa200870 Pgi genom
C	35	17.4	79.1	56520	2	AAZ01022	Aa201022 Wild type
C	36	17.2	78.2	286	2	AAV89814	Aa89814 EST clone
C	37	17.2	78.2	493	6	AB180978	Ab180978 Human ova
C	38	17.2	78.2	563	4	AA116914	Aa116914 Probe #68
C	39	17.2	78.2	563	4	ABA60895	Ab60895 Human foe
C	40	17.2	78.2	563	4	AA140790	Aa140790 Probe #94
C	41	17.2	78.2	563	4	ABA28878	Ab28878 Probe #73
C	42	17.2	78.2	563	4	AAK35074	Aa35074 Human bon
C	43	17.2	78.2	563	4	AAK09185	Aa09185 Human bra
C	44	17.2	78.2	563	4	ABS34826	Ab34826 Human liv
C	45	17.2	78.2	563	6	ABS09542	Ab09542 Human gen

ALIGNMENTS

RESULT 1	AAI41111	standard; DNA; 722 BP.
ID	AAI41111	
AC	AAI41111	
XX		
XX		
DT	16-OCT-2002	(first entry)
XX		
DB	GAML related Y-chromosome DNA sequence.	
XX		
KW	Goat embryo sexual identification technique; goat amelogenin gene; GAML;	
KW	sex-specific; gene; ds; Y-chromosome.	
XX		
OS	Capra hircus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	35..658
FT		/*tag= a
XX		/product= "Y-chromosome protein"
XX		
PN	TW454013-A.	
XX		
PD	11-SEP-2001.	
XX		
PF	10-NOV-1999;	99TW-00119616.
XX		
PR	10-NOV-1999;	99TW-00119616.
XX		
PA	(CHEN/) CHEN C.	
PA	(JANG/) JANG J.	
PA	(WENG/) WENG T.	
PA	(JENG/) JENG D.	
XX		
PI	Chen C, Jang J, Weng T, Jeng D;	
XX		
DR	WPI; 2002-442016/47.	
XX		
DR	P-PSDB; AAO22534.	
XX		
PT	Sex-specific sequence of goat amelogenin gene, useful for embryo sexual	
PT	identification, comprises high sensitivity even using single white blood	
XX	cell or cleavage c.	
XX		
PS	Disclosure; Page 28; 35pp; Chinese.	
XX		

CC The invention relates to a goat embryo sexual identification technique
CC with high efficiency, sensitivity and repeatability. This technique
CC involves separately cloning and sequencing the coding regions and the
CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
CC results indicate that there are sex-specific sequences in the fifth
CC intron of the gene. The major characteristics according to the present
CC invention include high sensitivity, applicable in sex identification even
CC only using a single white blood cell or a single cleavage cell of
CC blastula; high diagnostic efficiency, capable of identifying hundreds of
CC goat embryo in 3 hours; simple operation procedures without complicated
CC steps of DNA extraction and need no additional control group intron; and
CC can be applied on different species of goats. This polynucleotide
CC sequence represents a GAML related Y-chromosome DNA sequence of the
CC invention
XX

SO Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 6; Length 722;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22
|||||
46 GATTTGTTGGCTGCTCTG 67

Db

RESULT 2
AAL41110
ID AAL41110 standard; DNA; 752 BP.
XX
AC AAL41110;
XX
DT 16-OCT-2002 (first entry)
XX
DE GAML related X-chromosome DNA sequence.
XX
KW Goat embryo sexual identification technique; goat amelogenin gene; GAML;
XX sex-specific; gene; db; X-chromosome.
XX
OS Capra hircus.
XX
FH Key Location/Qualifiers
FT CDS 35..658
FT /*tag= a
FT /product= "X-chromosome protein"
XX
PN TW454013-A.
XX
PD 11-SEP-2001.
XX
PF 10-NOV-1999; 99TW-00119616.
XX
PR 10-NOV-1999; 99TW-00119616.
XX
PA (CHEN/) CHEN C.
PA (JANG/) JANG J.
PA (WENG/) WENG T.
PA (JENG/) JENG D.
XX
PI Chen C, Jang J, Weng T, Jeng D;
XX
DR WPI: 2002-442016/47.
DR P-PsDB; AAO22534.
XX
PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
PT identification, comprises high sensitivity even using single white blood
PT cell or cleavage c.
XX
PS Disclosure; Page 28; 35pp; Chinese.
XX
CC The invention relates to a goat embryo sexual identification technique
CC with high efficiency, sensitivity and repeatability. This technique
CC involves separately cloning and sequencing the coding regions and the

CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
CC results indicate that there are sex-specific sequences in the fifth
CC intron of the gene. The major characteristics according to the present
CC invention include high sensitivity, applicable in sex identification even
CC only using a single white blood cell or a single cleavage cell of
CC blastula; high diagnostic efficiency, capable of identifying hundreds of
CC goat embryo in 3 hours; simple operation procedures without complicated
CC steps of DNA extraction and need no additional control group intron; and
CC can be applied on different species of goats. This polynucleotide
CC sequence represents a GAML related X-chromosome DNA sequence of the
CC invention
XX

SO Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 6; Length 752;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22
|||||
46 GATTTGTTGGCTGCTCTG 67

Db

RESULT 3
ABN28389
ID ABN28389 standard; DNA; 65 BP.
XX
AC ABN28389;
XX
DT 15-JUL-2002 (first entry)
XX
DE Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Rattus norvegicus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB001903.
XX
PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI: 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 1137; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridizing selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular

CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 6; Length 65;
Best Local Similarity 95.5%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCCTG 22
Db 6 GATCTTGTTCCTGCTCCTG 27

RESULT 4

ADM80835
ID ADM80835 standard; cDNA; 270 BP.

AC ADM80835;

DT 03-JUN-2004 (first entry)

DE Human CADECM-22 encoding cDNA SEQ ID NO:64.

XX human; cell adhesion and extracellular matrix protein; CADECM;
XX neuroprotective; cytoskeletal; anorectic; immune disorder;
XX neurological disorder; developmental disorder;
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX Tangle disease; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 35..212
XX FT /*tag= a
XX FT /product= "CADECM-22"

XX WO2004015396-A2.

XX 19-FEB-2004.

XX 12-AUG-2003; 2003WO-US025418.

XX 13-AUG-2002; 2002US-0403781P.

XX 30-AUG-2002; 2002US-0407034P.

XX 13-SEP-2002; 2002US-0410566P.

XX 24-SEP-2002; 2002US-0413482P.

XX 25-SEP-2002; 2002US-0413890P.

XX 08-NOV-2002; 2002US-0424504P.

XX 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J;
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX Wang JT, Chien D, Yang YG;

XX MPI; 2004-191795/18.

XX P-PSDB; ADM80793.

XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.

PS Claim 5; SEQ ID NO 64; 272pp; English.

XX The present sequence encodes a human cell adhesion and extracellular
XX matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX cytoskeletal and anorectic activities. The CADECM polypeptides and
XX polynucleotides are useful in diagnosing, treating and preventing immune,
XX neurological, developmental, connective tissue and cell proliferative
XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX cancer, obesity and Tangle disease.

XX Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 12; Length 270;
Best Local Similarity 95.5%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTCCTG 22
Db 47 GATTTATTTGCTGCTCCTG 68

RESULT 5

ADM80836
ID ADM80836 standard; cDNA; 318 BP.

AC ADM80836;

DT 03-JUN-2004 (first entry)

DE Human CADECM-23 encoding cDNA SEQ ID NO:65.

XX human; cell adhesion and extracellular matrix protein; CADECM;
XX neuroprotective; cytoskeletal; anorectic; immune disorder;
XX neurological disorder; developmental disorder;
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX Tangle disease; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 35..260
XX FT /*tag= a
XX FT /product= "CADECM-23"

XX WO2004015396-A2.

XX 19-FEB-2004.

XX 12-AUG-2003; 2003WO-US025418.

XX 13-AUG-2002; 2002US-0403781P.

XX 30-AUG-2002; 2002US-0407034P.

XX 13-SEP-2002; 2002US-0410566P.

XX 24-SEP-2002; 2002US-0413482P.

XX 25-SEP-2002; 2002US-0413890P.

XX 08-NOV-2002; 2002US-0424504P.

XX 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J;
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX Wang JT, Chien D, Yang YG;

XX MPI; 2004-191795/18.

XX P-PSDB; ADM80794.

XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.
XX Claim 5; SEQ ID NO 65; 272pp; English.

```
XX The present sequence encodes a human cell adhesion and extracellular
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
CC cyostatic and anorectic activities. The CADECM polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing immune,
CC neurological, developmental, connective tissue and cell proliferative
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and Tangle disease.
XX
SQ Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 12; Length 318;
Best Local Similarity 95.5%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GATTTGTTGGCCCTCCTG 22
Db 47 GATTTGTTGGCCCTCCTG 68
RESULT 6
AAZ50832
ID AAZ50832 standard; DNA; 476 BP.
XX
AC AAZ50832;
XX
DT 31-MAY-2000 (first entry)
XX
DE Rat amelogenin gene (A4).
XX
KM Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;
KM chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;
KM extracellular matrix protein; tooth enamel; enamel mineralisation;
KM ameloblast; bone regeneration; composite cell construct; ds.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT exon 1..36
FT /*tag= a
FT /partial
FT /number= 1
FT exon 37..101
FT /*tag= b
FT /number= 2
FT CDS 48..317
FT /*tag= c
FT /product= "Rat amelogenin protein"
FT /transl_except= (pos:69..71, aa:Gly)
FT sig_peptide 48..95
FT /*tag= d
FT mat_peptide 96..314
FT /*tag= e
FT /label= Mature_rat_amelogenin_protein
FT exon 102..149
FT /*tag= f
FT /number= 3
FT exon 150..191
FT /*tag= g
FT /number= 4
FT exon 192..236
FT /*tag= h
FT /number= 5
FT exon 237..311
FT /*tag= i
FT /number= 6d
FT /note= "Comprises of gene segments 6a, b, c and d"
FT exon 312..317
FT /*tag= j
FT /number= 7
FT /note= "Includes the stop codon"
XX
PN WO200006734-A1.
```

```
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US017342.
XX
PR 29-JUL-1998; 98US-0094489P.
XX
PA (NOUN ) UNIV NORTHWESTERN.
XX
PI Vels A, Nebgen DR;
XX
DR WPI; 2000-205464/18.
XX
DR P-PDB; AAY45074.
XX
PT Novel amelogenin polypeptides and polynucleotides, useful for enhancing
PT bone generation in mammals and synthesizing bone matrix or articular
PT surfaces at implant sites.
XX
PS Example 2; Fig 11B; 79p; English.
XX
CC The present DNA sequence is the full-length rat amelogenin gene (A4),
CC comprising exons 1-7, including the exon segment 6d. It is derived from
CC the rat incisor odontoblast-pulp cDNA library. The splice variants of
CC this gene functions as an osteogenic or chondrogenic inducing molecule
CC (CIM), which is useful for enhancing bone or cartilage growth. It has
CC osteopathic activity. Amelogenin belongs to the family of extracellular
CC matrix proteins, in developing tooth enamel, that are produced by the
CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or
CC osteogenic inducing amelogenin molecules are useful to induce
CC differentiation of cells to the osteogenic and chondrogenic phenotypes
CC and can be used in a composite cell construct for bone and cartilage
CC regeneration. The polynucleotides can be employed to produce the
CC polypeptides by recombinant techniques
XX
SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 3; Length 476;
Best Local Similarity 95.5%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GATTTGTTGGCCCTCCTG 22
Db 59 GATCTGTTGGCCCTCCTG 80
RESULT 7
ADM80826
ID ADM80826 standard; cDNA; 549 BP.
XX
AC ADM80826;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human CADECM-13 encoding cDNA SEQ ID NO:55.
XX
KW human; cell adhesion and extracellular matrix protein; CADECM;
KW neuroprotective; cyostatic; anorectic; immune disorder;
KW neurological disorder; developmental disorder;
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW Tangle disease; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 36..491
FT /*tag= a
FT /product= "CADECM-13"
XX
PN WO2004015396-A2.
XX
PD 19-FEB-2004.
XX
PF 12-AUG-2003; 2003WO-US025418.
```



```

XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424904P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX PA (INCY-) INCYTE CORP.
XX PI Eliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margula JP, Swarnakar A, Chawla NK, Ramkumar J;
XX PI Hafalla AJM, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;
XX DR WPI, 2004-191795/18.
XX P-PsDB; ADM80784.
XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.
XX PS Claim 5; SEQ ID NO 55; 272pp; English.
XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADPCM. CADPCM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADPCM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative
XX CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX CC cancer, obesity and Tangle disease.
XX SQ Sequence 549 BP, 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
XX
Query Match 92.7%; Score 20.4; DB 12; Length 549;
Best Local Similarity 95.5%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GATTTGTTGGCTGCTCTG 22
Db 47 GATTTATTTGGCTGCTCTG 68

```

RESULT 8
ADBS9026
ID ADBS9026 standard; DNA; 556 BP.
AC ADBS9026;
XX
DT 04-DEC-2003 (first entry)
XX
DE Toxicity-related gene, SEQ ID 4052.
XX
KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; de.
XX
OS Unidentified.
XX
PN WO2003064624-A2.
XX
PD 07-AUG-2003.
XX
PF 31-JAN-2003; 2003WO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
XX
PR 15-MAR-2002; 2002US-0364045P.
XX
PR 15-MAR-2002; 2002US-0364055P.
XX
PR 30-DEC-2002; 2002US-0436643P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

```

XX DR WPI, 2003-689530/65.
XX PR Predicting a toxic effect of a compound, useful in identifying toxicity
XX PR markers in liver tissues or cells for drug screening and toxicity assays,
XX PR comprises preparing gene expression profile of tissue or cells exposed to
XX PR the compound.
XX PS Claim 1; SEQ ID NO 4052; 1156pp; English.
XX
XX CC The present invention relates to a method for predicting a toxic effect
XX CC of a compound. The method comprises preparing a gene expression profile
XX CC of a tissue or cell sample exposed to the compound, and comparing the
XX CC gene expression profile to a database comprising SEQ ID 1-4925, where
XX CC differential expression of the gene indicates at least one toxic effect.
XX CC The method is useful for predicting at least one toxic effect of a
XX CC compound, predicting hepatotoxicity or the progression of a toxic effect
XX CC of a compound, identifying an agent that modulates the onset or
XX CC progression of a toxic response, predicting the cellular pathways that a
XX CC compound modulates in a cell, and identifying an agent that modulates at
XX CC least one activity of a protein. The method and compositions of the
XX CC present invention using a database of genes having liver toxin-induced
XX CC differential expression, are useful in identifying toxicity markers in
XX CC liver tissues or cells for drug screening and toxicity assays. Note: The
XX CC sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
XX
Query Match 92.7%; Score 20.4; DB 10; Length 556;
Best Local Similarity 95.5%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GATTTGTTGGCTGCTCTG 22
Db 33 GATCTTGTGGCTGCTCTG 54

```

RESULT 9
ADBS3782
ID ADBS3782 standard; DNA; 556 BP.
XX
AC ADBS3782;
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; de.
XX
OS Rattus norvegicus.
XX
PN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003WO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
XX
PR 13-MAR-2002; 2002US-0363534P.
XX
PR 08-APR-2002; 2002US-0370248P.
XX
PR 10-APR-2002; 2002US-0371134P.
XX
PR 10-APR-2002; 2002US-0371135P.
XX
PR 10-APR-2002; 2002US-0371150P.
XX
PR 11-APR-2002; 2002US-0371413P.
XX
PR 11-APR-2002; 2002US-0373601P.
XX
PR 19-APR-2002; 2002US-0374139P.
XX
PR 22-APR-2002; 2002US-0376370P.
XX
PR 08-MAY-2002; 2002US-0378370P.
XX
PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394238P.
PR 09-JUL-2002; 2002US-0394239P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
PA (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Potter M, Johnson K, Higgs B, Castle A, Orr M,
PI Elshoff M;
XX
XX WPI; 2003-731472/69.
XX
XX Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
XX Claim 44; SEQ ID NO 4324; 874bp; English.
XX
XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
XX SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
XX
XX Query Match 92.7%; Score 20.4; DB 10; Length 556;
XX Best Local Similarity 95.5%; Pred. No. 18;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GATTTGTTGCTGCTGCTG 22
XX |||||
XX 33 GATCTGTGTGCTGCTGCTG 54
XX
XX RESULT 10
XX ADM80827
XX ID ADM80827 standard; cDNA; 623 BP.
XX
XX ADM80827;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human CADECM-14 encoding cDNA SEQ ID NO:56.
XX
XX human; cell adhesion and extracellular matrix protein; CADECM;
XX neuroprotective; cytostatic; anorectic; immune disorder;
XX neurological disorder; developmental disorder;
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX Tangier disease; gene; ss.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 36..563
XX FT /*tag= a
XX FT /product= "CADECM-14"
XX
XX WO2004015396-A2.
XX
XX 19-FEB-2004.
XX
XX PD

XX
XX 12-AUG-2003; 2003WO-US025418.
XX
XX 13-AUG-2002; 2002US-0403781P.
XX 10-AUG-2002; 2002US-0407034P.
XX 13-SEP-2002; 2002US-0410566P.
XX 24-SEP-2002; 2002US-0413482P.
XX 25-SEP-2002; 2002US-0413890P.
XX 08-NOV-2002; 2002US-0424904P.
XX 13-NOV-2002; 2002US-0426222P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
XX Hafalia AFA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake UT;
XX Wang JT, Chien D, Yang YG;
XX
XX WPI; 2004-191795/18.
XX
XX P-PSDB; ADM80785.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.
XX
XX Claim 5; SEQ ID NO 56; 272bp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
XX matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX cytostatic and anorectic activities. The CADECM polypeptides and
XX polynucleotides are useful in diagnosing, treating and preventing immune,
XX neurological, developmental, connective tissue and cell proliferative
XX disorders including cancer; e.g. breast, prostate, ovarian, lung or colon
XX cancer, obesity and Tangier disease.
XX
XX SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
XX
XX Query Match 92.7%; Score 20.4; DB 12; Length 623;
XX Best Local Similarity 95.5%; Pred. No. 18;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GATTTGTTGCTGCTGCTG 22
XX |||||
XX 47 GATTTATTTGCTGCTGCTG 68
XX
XX RESULT 11
XX AAZ07020
XX ID AAZ07020 standard; DNA; 750 BP.
XX
XX AAZ07020;
XX
XX 15-NOV-1999 (first entry)
XX
XX Amelogenin X nucleotide sequence.
XX
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
XX multiple allelic sites; apolipoprotein E; apoE; coronary artery disease;
XX Alzheimer's disease; ds.
XX
XX Unidentified.
XX
XX WO9940226-A2.
XX
XX 12-AUG-1999.
XX
XX 08-JAN-1999; 99WO-US000499.
XX
XX 04-FEB-1998; 98US-00018595.
XX
XX (PEKE) PERKIN-ELMER CORP.
XX
XX Liivak KI, Goodsaiaid F;
XX
XX PI

```
XX DR WPI; 1999-539985/45.
XX PT 5' nuclease amplification assay using fluorescence-quencher probes for
XX PT determination of a genotype at multiple allelic sites.
XX PS Disclosure; Fig 10; 95pp; English.
XX CC The present invention describes first and second sets of fluorescer-
XX CC quencher probes used simultaneously in a 5' nuclease assay to identify
XX CC which members of a first or second set of substantially homologous
XX CC sequences are present in a DNA sample. The method can be used to genotype
XX CC a sample of genomic DNA at two or more different allelic sites.
XX CC Generating a fluorescence spectrum and signature for each genotype, which
XX CC uniquely reflects the assay's inherent inefficiency for that genotype
XX CC given the particular conditions, probes and primers used, the genotype of
XX CC unknown sequences can be determined. The assay was shown to be useful for
XX CC determining apog genotypes. The assay can be used as a diagnostic tool
XX CC for assessing the risk for coronary artery disease and/or late-onset
XX CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
XX CC possible to determine a genotype at two or more allelic sites in a single
XX CC reaction. This approach is much faster than previous approaches to
XX CC genotyping genes having two or more allelic sites, such as the
XX CC apolipoprotein E gene. A key advantage of the method for determining the
XX CC genotype of a sample of DNA at multiple allelic sites is that it does not
XX CC rely on 5' nuclease assay working with 100% efficiency to distinguish
XX CC between substantially homologous sequences such as alleles. The present
XX CC sequence represent the nucleotide sequence for amelogenin X, which is
XX CC used in the exemplification of the present invention
SQ Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Query Match          92.7%; Score 20.4; DB 2; Length 750;
Best Local Similarity 95.5%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTG 22
   |||||
DB 80 GATTTATTGCTGCTGCTG 101

RESULT 12
AAZ07018
ID AAZ07018 standard; DNA; 793 BP.
XX AC AAZ07018;
XX DT 15-NOV-1999 (first entry)
XX DE Amelogenin X nucleotide sequence.
XX KW Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
XX KW multiple allelic site; apolipoprotein E; apog; coronary artery disease;
XX KW Alzheimer's disease; ss.
XX OS Unidentified.
XX PN WO940226-A2.
XX PD 12-AUG-1999.
XX PF 08-JAN-1999; 99WO-US000499.
XX PR 04-FEB-1998; 98US-00018595.
XX PA (PERKE ) PERKIN-ELMER CORP.
XX PI Ljvak KJ, Goodsaiaid F;
XX DR WPI; 1999-539985/45.
XX PT 5' nuclease amplification assay using fluorescence-quencher probes for
XX PT determination of a genotype at multiple allelic sites.
```

```
XX PS Disclosure; Fig 8A; 95pp; English.
XX CC The present invention describes first and second sets of fluorescer-
XX CC quencher probes used simultaneously in a 5' nuclease assay to identify
XX CC which members of a first or second set of substantially homologous
XX CC sequences are present in a DNA sample. The method can be used to genotype
XX CC a sample of genomic DNA at two or more different allelic sites.
XX CC Generating a fluorescence spectrum and signature for each genotype, which
XX CC uniquely reflects the assay's inherent inefficiency for that genotype
XX CC given the particular conditions, probes and primers used, the genotype of
XX CC unknown sequences can be determined. The assay was shown to be useful for
XX CC determining apog genotypes. The assay can be used as a diagnostic tool
XX CC for assessing the risk for coronary artery disease and/or late-onset
XX CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
XX CC possible to determine a genotype at two or more allelic sites in a single
XX CC reaction. This approach is much faster than previous approaches to
XX CC genotyping genes having two or more allelic sites, such as the
XX CC apolipoprotein E gene. A key advantage of the method for determining the
XX CC genotype of a sample of DNA at multiple allelic sites is that it does not
XX CC rely on 5' nuclease assay working with 100% efficiency to distinguish
XX CC between substantially homologous sequences such as alleles. The present
XX CC sequence represent the nucleotide sequence for amelogenin X, which is
XX CC used in the exemplification of the present invention
SQ Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match          92.7%; Score 20.4; DB 2; Length 793;
Best Local Similarity 95.5%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTG 22
   |||||
DB 80 GATTTATTGCTGCTGCTG 101

RESULT 13
ADQ18492
ID ADQ18492 standard; DNA; 259202 BP.
XX AC ADQ18492;
XX DT 26-AUG-2004 (first entry)
XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.
XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX KW ds.
XX OS Homo sapiens.
XX PN WO2004048938-A2.
XX PD 10-JUN-2004.
XX PF 26-NOV-2003; 2003WO-US038193.
XX PR 26-NOV-2002; 2002US-0429739P.
XX PA (PROT-) PROTEIN DESIGN LABS INC.
XX PI Aziz N, Ginsburg WM, Zlotnick A;
XX DR WPI; 2004-441208/41.
XX PT Early detection of soft tissue sarcoma comprises determining expression
XX PT of a gene in a first soft tissue sample and a normal soft tissue sample
XX PT and comparing the gene expression, also useful in treating soft tissue
XX PT sarcoma.
XX PS Example 2; SEQ ID NO 1311; 210pp; English.
XX CC The invention relates to a novel method for detecting soft tissue sarcoma
```

CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

SQ Sequence 259202 BP; 81699 A; 51421 G; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 12; Length 259202;
 Best Local Similarity 95.5%; Pred. No. 39;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGCTG 22
 |||||
 Db 69884 GATTTATTCCTGCTGCTGCTG 69905

RESULT 14
 AAX18413
 ID AAX18413 standard; DNA; 20 BP.

AC AAX18413;

DT 11-MAY-1999 (first entry)

DE PCR primer BE2(+) for bovine amelogenin gene.

KW Amelogenin gene; cow sexing; Holstein dairy cow; bAML intron 5;
 bML bovine embryo sexing; PCR primer; 89.

OS Synthetic.
 OS Bos sp.

PN US5876942-A.

PD 02-MAR-1999.

PF 24-JUL-1997; 97US-00899811.

PR 24-JUL-1997; 97US-00899811.

PA (NASC-) NAT SCI COUNCIL REPUBLIC OF CHINA.

PI Choo K, Wang C, Cheng WT, Chen C, Hu C;

DR WPI; 1999-189629/16.

PT New oligonucleotide primers based on bovine amelogenin gene, intron 5
 PT sequences - useful for sexing cows by Polymerase Chain Reaction studies.

PS Disclosure; Col 6; 28pp; English.

CC This sequence is a PCR primer for the Holstein cow amelogenin (bAML) gene
 CC The invention relates to an oligonucleotide primer set, useful for bovine
 CC embryo sexing, that comprises two primers, each of which can hybridise
 CC specifically and simultaneously, to an intron 5 sequence of bAML, located
 CC on the bovine X and Y chromosomes. The primers may be used in a rapid,
 CC highly reproducible and sensitive method for determining the sex of
 CC bovine embryos, which involves PCR of the bAML genes located on the X and
 CC Y chromosomes of Holstein dairy cattle. In order to use PCR in sex-
 CC determination studies, a nucleotide sequence, specific against sex, has
 CC to be produced (e.g. one associated with testis determining factor).
 CC However, in this PCR based method, each primer can only recognise DNA
 CC fragments from one, not both, of the sex chromosomes, therefore, internal
 CC control primers, derived from the subject gene have to be added to the
 CC reaction. This can result in competition between the primers, or the

CC formation of dimer primers during amplification, rendering the results
 CC inaccurate. The primers overcome this problem, as they are homologous to
 CC both the X and Y chromosomes, and so amplify DNA from both chromosomes
 CC simultaneously, allowing gender to be determined by quick, simple and
 CC accurate PCR and electrophoresis

SQ Sequence 20 BP; 1 A; 6 C; 3 G; 10 T; 0 U; 0 Other;

Query Match 90.9%; Score 20; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ATTTGTTGCTGCTGCTGCT 21
 |||||
 Db 1 ATTTGTTGCTGCTGCTGCT 20

RESULT 15
 ABK92154
 ID ABK92154 standard; DNA; 2190 BP.

AC ABK92154;

DT 15-AUG-2002 (first entry)

DE Prostate cancer-associated DNA sequence #40.

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
 KW gene therapy; gene; ds.

OS Mammalia.

PN WO200230268-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US032045.

PR 13-OCT-2000; 2000US-00687576.

PR 08-DEC-2000; 2000US-00733288.

PR 08-DEC-2000; 2000US-00733742.

PR 24-JAN-2001; 2001US-0263957P.

PR 16-MAR-2001; 2001US-0276791P.

PR 16-MAR-2001; 2001US-0276888P.

PR 06-APR-2001; 2001US-0281922P.

PR 24-APR-2001; 2001US-0286214P.

PR 30-APR-2001; 2001US-00847046.

PR 04-MAY-2001; 2001US-0288589P.

PA (EOSB-) EOS BIOTECHNOLOGY INC.

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevez P;

DR WPI; 2002-471335/50.

DR P-PSDB; ABG61839.

PT Detecting a prostate cancer-associated transcript in a cell in a patient,
 PT by determining if prostate cancer-associated genes are expressed in a
 PT prostate tissue.

PS Claim 22; Page 331; 436pp; English.

CC The present invention relates to methods of detecting a prostate cancer-
 CC associated transcript in a cell from a patient. The method comprises
 CC contacting a biological sample from the patient with prostate cancer-
 CC associated polynucleotides (designated PC genes) that selectively
 CC hybridise to a sequence that is at least 80% identical to them. The
 CC prostate cancer-associated polynucleotide sequences are differentially
 CC expressed in prostate tumour tissue or in prostate cancer and are derived
 CC from the tissues of various organisms such as humans or other mammals
 CC (e.g. mice, sheep and dogs). The methods of the invention are useful for
 CC diagnosing and treating prostate cancer in mammals. The prostate cancer-

CC associated genes are useful for diagnosing or treating prostate cancer,
CC as well as for identifying modulators of prostate cancer or agents that
CC inhibit prostate cancer. The nucleic acid sequences are particularly
CC useful in gene therapy, as a vaccine or in antisense applications.
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
CC sequences

XX
SQ Sequence 2190 BP; 583 A; 488 C; 494 G; 625 T; 0 U; 0 Other;

Query Match 88.2%; Score 19.4; DB 6; Length 2190;

Best Local Similarity 95.2%; Pred. No. 60;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATTTGTTGGCTGCTCTCG 22

Db 2026 ATTTGTTGGCTGCTCTCG 2046

Search completed: February 8, 2005, 13:56:51
Job time : 35.1839 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 8, 2005, 13:03:55 : Search time 5.94179 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22
Sequence: 1 gatctgtgtgctgctgctctg 22

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.4	92.7	476	4 US-09-744-128-3	Sequence 3, Appl1
2	20.4	92.7	500	2 US-09-018-595B-3	Sequence 3, Appl1
3	20.4	92.7	500	3 US-09-324-709A-3	Sequence 3, Appl1
4	20.4	92.7	793	2 US-018-595B-1	Sequence 1, Appl1
5	20.4	92.7	793	3 US-09-324-709A-1	Sequence 5, Appl1
6	20	90.9	20	2 US-08-899-811-5	Sequence 1, Appl1
7	19.4	88.2	4181	4 US-09-976-594-366	Sequence 366, App
8	19	86.4	19	2 US-08-899-811-4	Sequence 4, Appl1
9	18.8	85.5	802	2 US-09-018-595B-2	Sequence 2, Appl1
10	18.8	85.5	802	3 US-09-324-709A-2	Sequence 35462, A
11	17.8	80.9	173	4 US-09-513-999C-35462	Sequence 110, App
12	17.4	79.1	116	3 US-09-338-907-110	Sequence 110, App
13	17.4	79.1	116	3 US-09-318-207-110	Sequence 124, App
14	17.4	79.1	5326	3 US-09-338-907-124	Sequence 124, App
15	17.4	79.1	5326	3 US-09-218-207-124	Sequence 124, App
16	17.4	79.1	56516	2 US-08-996-306-1	Sequence 1, Appl1
17	17.4	79.1	56516	3 US-09-338-907-1	Sequence 1, Appl1
18	17.4	79.1	56516	3 US-09-218-207-1	Sequence 1, Appl1
19	17.4	79.1	56520	3 US-09-338-907-179	Sequence 179, App
20	17.4	79.1	56520	3 US-09-218-207-179	Sequence 179, App
21	16.8	76.4	40	4 US-09-313-221A-40	Sequence 40, Appl1
22	16.8	76.4	2734	4 US-09-461-657B-13	Sequence 13, Appl1
23	16.8	76.4	2734	4 US-09-461-657B-11	Sequence 11, Appl1
24	16.8	76.4	2912	4 US-09-461-657B-12	Sequence 12, Appl1
25	16.8	76.4	2987	3 US-07-757-342D-1	Sequence 1, Appl1
26	16.8	76.4	2987	4 US-09-461-657B-1	Sequence 1, Appl1
27	16.8	76.4	3018	4 US-09-016-434-1468	Sequence 1468, Ap

C	28	16.4	74.5	2438	6	5432081-1	Patent No. 5432081
	29	16.2	73.6	433	4	US-09-621-976-13460	Sequence 13460, A
	30	16.2	73.6	586	4	US-09-669-751-55	Sequence 55, Appl1
	31	16.2	73.6	2148	3	US-09-219-983A-19	Sequence 19, Appl1
	32	16.2	73.6	2492	3	US-09-219-983A-18	Sequence 18, Appl1
	33	16.2	73.6	5852	4	US-09-853-768-10	Sequence 10, Appl1
	34	16.2	73.6	7037	4	US-09-853-768-3	Sequence 3, Appl1
	35	15.8	71.8	282	4	US-09-248-796A-7539	Sequence 7539, Ap
	36	15.8	71.8	619	4	US-09-270-767-12644	Sequence 12644, A
	37	15.8	71.8	1407	3	US-09-505-250-2	Sequence 2, Appl1
	38	15.8	71.8	1917	3	US-08-808-346-1	Sequence 1, Appl1
	39	15.8	71.8	1926	1	US-07-901-703-12	Sequence 12, Appl1
	40	15.8	71.8	1926	1	US-08-147-023-26	Sequence 26, Appl1
	41	15.8	71.8	1926	1	US-08-278-729A-22	Sequence 22, Appl1
	42	15.8	71.8	1926	1	US-08-480-528A-9	Sequence 9, Appl1
	43	15.8	71.8	1926	1	US-08-479-666-9	Sequence 9, Appl1
	44	15.8	71.8	1926	1	US-08-155-343A-22	Sequence 22, Appl1
	45	15.8	71.8	1926	1	US-08-406-672-22	Sequence 22, Appl1

ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3

Query Match          92.7%; Score 20.4; DB 4; Length 476;
Best Local Similarity 95.5%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTCTG 22
      ||| ||||| ||||| ||||| |||||
Db      59 GATCTGTGCTGCTGCTCTG 80

RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Weitz,
; ADDRESSER: Wilson Somsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-018-595B-3

Query Match 92.7%; Score 20.4; DB 2; Length 500;
Best Local Similarity 95.5%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTG 22
Db 80 GATTTATTGCTGCTGCTG 101

RESULT 3
US-09-324-709A-3
Sequence 3, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-324-709A-3

Query Match 92.7%; Score 20.4; DB 3; Length 500;
Best Local Similarity 95.5%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTG 22
Db 80 GATTTATTGCTGCTGCTG 101

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

Query Match 92.7%; Score 20.4; DB 2; Length 793;
Best Local Similarity 95.5%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTG 22
Db 80 GATTTATTGCTGCTGCTG 101

RESULT 5
US-09-324-709A-1
; Sequence 1, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: David J. Weitz,
; ADDRESSES: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-324-709A-1
Query Match 92.7%; Score 20.4; DB 3; Length 793;
Best Local Similarity 95.5%; Pred. No. 1.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GATTGTGCTGCTGCTCTG 22
DB 80 GATTGTGCTGCTGCTCTG 101
RESULT 6
US-08-899-811-5
; Sequence 5, Application US/08899811
; Patent No. 5876942
; GENERAL INFORMATION:
; APPLICANT: CHENG, WINSTON T.K.
; APPLICANT: CHOO, KONG-BUNG
; APPLICANT: HU, CHE-LIN
; APPLICANT: WANG, CHIH-HUA
; APPLICANT: CHEN, CHUAN-WU
; TITLE OF INVENTION: A PROCESS FOR SEXING COW EMBRYOS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: BACON & THOMAS
; STREET: 625 SLATERS LANE - FOURTH FLOOR
; CITY: ALEXANDRIA

STATE: VIRGINIA
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,811
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FICHTER, RICHARD E.
REGISTRATION NUMBER: 26,382
REFERENCE/DOCKET NUMBER: REF/CHEN/881
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-683-0500
TELEFAX: 703-683-1080
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-899-811-5
Query Match 90.9%; Score 20; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATTGTGTTGCTGCTGCTCT 21
DB 1 ATTGTGTTGCTGCTGCTCT 20
RESULT 7
US-09-976-594-366
; Sequence 366, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jemmy
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 366
; LENGTH: 4181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 337156.2
US-09-976-594-366
Query Match 88.2%; Score 19.4; DB 4; Length 4181;
Best Local Similarity 95.2%; Pred. No. 6.3;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ATTGTGTTGCTGCTGCTCTG 22
DB 2056 ATTGTGTTGCTGCTGCTCTG 2076
RESULT 8
US-08-899-811-4/c
; Sequence 4, Application US/08899811

RESULT 9
 US-09-018-595B-2
 Sequence 2, Application US/09018595B
 Patent No. 596223
 GENERAL INFORMATION:
 APPLICANT: Perkin-Elmer Corporation,
 APPLICANT: Applied Biosystems Division
 TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
 TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David J. Weitz,
 ADDRESSEE: Wilson Sonsini Goodrich & Rosati
 STREET: 650 Page Mill Road
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1050
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: Microsoft Windows 95
 SOFTWARE: Wordperfect for Windows 6.0,

```

RESULT 10
US-09-324-709A-2
: Sequence 2, Application US/09324709A
: Patent No. 6134707
: GENERAL INFORMATION:
: APPLICANT: Perkin-Elmer Corporation,
: APPLICANT: Applied Biosystems Division
: TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
: TITLE OF INVENTION: SITES
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David J. Weitz,
: ADDRESSEE: Wilson Sonsini Goodrich & Rosati
: STREET: 650 Page Mill Road
: City: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304-1050
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM compatible
: OPERATING SYSTEM: Microsoft Windows 95
: SOFTWARE: Wordperfect for windows 6.0,
: SOFTWARE: ASCII (DOS) TEXT format
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/324,709A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: David J. Weitz
: REGISTRATION NUMBER: 38,362
: REFERENCE/DOCKET NUMBER: 16842-758
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 493-9300
: TELEFAX: (650) 493-6811
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 802 nucleotides

```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match 85.5%; Score 18.8; DB 3; Length 802;
Best Local Similarity 90.9%; Pred. No. 8.5;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22
DB 80 GATTTGTTGCTGCTGCTG 101

RESULT 11
US-09-513-999C-35462
Sequence 35462, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.Y.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59,US2,REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 35462
LENGTH: 173
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-35462

Query Match 80.9%; Score 17.8; DB 4; Length 173;
Best Local Similarity 90.5%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCT 21
DB 13 GATTTGTTGCTGCTCCT 33

RESULT 12
US-09-338-907-110
Sequence 110, Application US/09338907
Patent No. 6265546
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CP1P
CURRENT APPLICATION NUMBER: US/09/338,907
CURRENT FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 110
LENGTH: 116
TYPE: DNA
ORGANISM: Homo sapiens
US-09-338-907-110

Query Match 79.1%; Score 17.4; DB 3; Length 116;
Best Local Similarity 94.7%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCCTG 22
DB 72 TTGTGTTGCTGCTCCTG 90

RESULT 13
US-09-218-207-110
Sequence 110, Application US/09218207
Patent No. 6346381
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: Prostate cancer gene
FILE REFERENCE: GENSET.018CP1
CURRENT APPLICATION NUMBER: US/09/218,207
CURRENT FILING DATE: 1998-12-22
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 110
LENGTH: 116
TYPE: DNA
ORGANISM: Homo sapiens
US-09-218-207-110

Query Match 79.1%; Score 17.4; DB 3; Length 116;
Best Local Similarity 94.7%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCCTG 22
DB 72 TTGTGTTGCTGCTCCTG 90

RESULT 14
US-09-338-907-124
Sequence 124, Application US/09338907
Patent No. 6265546
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilya, Chumakov
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET.18CP1P
CURRENT APPLICATION NUMBER: US/09/338,907
CURRENT FILING DATE: 1999-06-23
EARLIER APPLICATION NUMBER: 08/996,306
EARLIER FILING DATE: 1997-12-22
EARLIER APPLICATION NUMBER: 60/099,658
EARLIER FILING DATE: 1998-09-09
EARLIER APPLICATION NUMBER: 09/218,207
EARLIER FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 124
LENGTH: 5324
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 31..588
NAME/KEY: polyA_signal
LOCATION: 5297..5302

/ OTHER INFORMATION: AATAAA
US-09-338-907-124

Query Match 79.1%; Score 17.4; DB 3; Length 5326;
Best Local Similarity 94.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTG 22
|||||
Db 657 TTGTTCCCTGCTCTG 675

RESULT 15
US-09-218-207-124
; Sequence 124; Application US/09218207
; Patent No. 6346381
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumentfeld, Marta
; APPLICANT: Ilya, Chumakov
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Prostate cancer gene
; FILE REFERENCE: GENSET.018CP1
; CURRENT APPLICATION NUMBER: US/09/218.207
; CURRENT FILING DATE: 1998-12-22
; EARLIER APPLICATION NUMBER: 08/996,306
; EARLIER FILING DATE: 1997-12-22
; EARLIER APPLICATION NUMBER: 60/099,658
; EARLIER FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 578
; SOFTWARE: Patent.pm
; SEQ ID NO 124
; LENGTH: 5324
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..588
; NAME/KEY: polyA_signal
; LOCATION: 5297..5302
; OTHER INFORMATION: AATAAA
US-09-218-207-124

Query Match 79.1%; Score 17.4; DB 3; Length 5326;
Best Local Similarity 94.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTGCTG 22
|||||
Db 657 TTGTTCCCTGCTCTG 675

Search completed: February 8, 2005, 22:25:50
Job time : 8.0529 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 16:37:21 ; Search time 29.6449 Seconds
(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22
Sequence: 1 gatttgcctgcctgcctcctg 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/us07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	92.7	65	US-09-908-975-1137	Sequence 1137, App
2	20.4	92.7	500	US-10-104-774-3	Sequence 3, Appl1
3	20.4	92.7	500	US-10-455-150-3	Sequence 3, Appl1
4	20.4	92.7	600	US-10-027-632-287049	Sequence 287049,
5	20.4	92.7	600	US-10-027-632-287050	Sequence 287050,
6	20.4	92.7	600	US-10-027-632-287049	Sequence 287049,
7	20.4	92.7	600	US-10-027-632-287050	Sequence 287050,
8	20.4	92.7	793	US-10-104-774-1	Sequence 1, Appl1
9	20.4	92.7	793	US-10-455-150-1	Sequence 1, Appl1
10	20.4	92.7	793	US-10-755-889-605	Sequence 605, App
11	20.4	92.7	259202	US-10-723-860-1311	Sequence 1311, App
12	19.4	88.2	228	US-10-674-124A-17105	Sequence 17105, A

13	19.4	88.2	2190	US-10-295-027-581	Sequence 581, App
14	19.4	88.2	2681	US-10-789-090-1	Sequence 1, Appl1
15	19.4	88.2	2685	US-09-964-824A-257	Sequence 257, App
16	19.4	88.2	2685	US-10-205-823-366	Sequence 366, App
17	19.4	88.2	2685	US-10-757-262-19	Sequence 19, Appl1
18	18.8	85.5	802	US-10-104-774-2	Sequence 2, Appl1
19	18.8	85.5	802	US-10-455-150-2	Sequence 2, Appl1
20	18.8	85.5	852	US-10-723-860-5797	Sequence 5797, Ap
21	18.4	83.6	1287	US-09-764-847-144	Sequence 144, App
22	18.4	83.6	1287	US-10-092-154-144	Sequence 144, App
23	18.4	83.6	1295	US-09-764-847-495	Sequence 495, App
24	18.4	83.6	1295	US-10-092-154-495	Sequence 495, App
25	18.4	83.6	168276	US-10-087-192-142	Sequence 142, App
26	18	81.8	668	US-10-027-632-34029	Sequence 34029, A
27	18	81.8	668	US-10-027-632-34029	Sequence 34029, A
28	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
29	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
30	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
31	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
32	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
33	17.8	80.9	598	US-10-027-632-217667	Sequence 217667,
34	17.4	79.1	80	US-10-384-245-267	Sequence 267, App
35	17.4	79.1	116	US-09-901-484A-110	Sequence 110, App
36	17.4	79.1	116	US-09-901-484A-110	Sequence 110, App
37	17.4	79.1	349	US-10-674-124A-26540	Sequence 26540, A
38	17.4	79.1	2756	US-10-062-674-16251	Sequence 16251, Ap
39	17.4	79.1	3923	US-10-152-319A-1931	Sequence 1931, Ap
40	17.4	79.1	5324	US-09-901-484A-124	Sequence 124, App
41	17.4	79.1	5324	US-09-901-484A-124	Sequence 124, App
42	17.4	79.1	56516	US-09-853-526-124	Sequence 1, Appl1
43	17.4	79.1	56516	US-09-853-526-1	Sequence 1, Appl1
44	17.4	79.1	56520	US-09-901-484A-179	Sequence 179, App
45	17.4	79.1	56520	US-09-853-526-179	Sequence 179, App

ALIGNMENTS

RESULT 1
US-09-908-975-1137
; Sequence 1137, Application US/0908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Shimon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 3668-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1137
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-908-975-1137

Query Match 92.7%; Score 20.4; DB 10; Length 65;

Best Local Similarity 95.5%; Pred. No. 9.6; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 6 GATTGTGTTGCTGCTGCTGCTG 22
GATCTGTGCTGCTGCTGCTGCTG 27

```
RESULT 2
US-10-104-774-3
; Sequence 3, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-3

Query Match          92.7%; Score 20.4; DB 13; Length 500;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGCTG 22
        ||||| ||||| ||||| |||||
Db      80 GATTTATTGCTGCTGCTGCTG 101

RESULT 3
US-10-455-150-3
; Sequence 3, Application US/10455150
; Publication No. US2004005302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
```

```
STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match          92.7%; Score 20.4; DB 16; Length 500;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGCTG 22
        ||||| ||||| ||||| |||||
Db      80 GATTTATTGCTGCTGCTGCTG 101

RESULT 4
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

Query Match          92.7%; Score 20.4; DB 13; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGCTG 22
        ||||| ||||| ||||| |||||
Db      245 GATTTATTGCTGCTGCTGCTG 224

RESULT 5
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827,129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
```

```

; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```

Query Match          92.7%; Score 20.4; DB 13; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 GATTTGTTTGCTGCTGCTG 22
    ||||| ||||| ||||| |||||
DB 245 GATTTATTGCTGCTGCTG 224
```

```

RESULT 6
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```

Query Match          92.7%; Score 20.4; DB 15; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 GATTTGTTTGCTGCTGCTG 22
    ||||| ||||| ||||| |||||
DB 245 GATTTATTGCTGCTGCTG 224
```

```

RESULT 7
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```

Query Match          92.7%; Score 20.4; DB 15; Length 600;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 GATTTGTTTGCTGCTGCTG 22
    ||||| ||||| ||||| |||||
DB 245 GATTTATTGCTGCTGCTG 224
```

```

RESULT 8
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
; FILE REFERENCE: 16842-782
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1
```

```

Query Match          92.7%; Score 20.4; DB 13; Length 793;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 GATTTGTTTGCTGCTGCTG 22
    ||||| ||||| ||||| |||||
DB 80 GATTTATTGCTGCTGCTG 101
```

```

RESULT 9
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
```

TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: David J. Weitz,
Wilson Sonsini Goodrich & Rosati
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match 92.7%; Score 20.4; DB 16; Length 793;
Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22
|||||
DB 80 GATTTATTGCTGCTCTG 101

RESULT 10
US-10-755-889-605
Sequence 605, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
PRIOR FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2
SEQ ID NO 605
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-605

Query Match 92.7%; Score 20.4; DB 17; Length 793;

Best Local Similarity 95.5%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22
|||||
DB 80 GATTTATTGCTGCTCTG 101

RESULT 11
US-10-723-860-1311
Sequence 1311, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Nacasha
APPLICANT: Ginsburg, Wendy M.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: Methods of diagnosis of Soft Tissue Sarcoma, Compositions &
FILE REFERENCE: 05892, 0193, NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
PRIOR FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1311
LENGTH: 259202
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 92.7%; Score 20.4; DB 18; Length 259202;
Best Local Similarity 95.5%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTG 22
|||||
DB 69884 GATTTATTGCTGCTCTG 69905

RESULT 12
US-10-674-124A-17105
Sequence 17105, Application US/10674124A
Publication No. US2004019797A1
GENERAL INFORMATION:
APPLICANT: INOKO, Hidetoshi
APPLICANT: TAMURA, Gen
TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
FILE REFERENCE: ORIN-003CIP
CURRENT APPLICATION NUMBER: US/10/674,124A
PRIOR FILING DATE: 2003-09-26
PRIOR APPLICATION NUMBER: 10/257,511
PRIOR FILING DATE: 2003-03-07
PRIOR APPLICATION NUMBER: PCT/JP00/07621
PRIOR FILING DATE: 2000-10-30
PRIOR APPLICATION NUMBER: JP2000-112699
PRIOR FILING DATE: 2000-04-13
PRIOR APPLICATION NUMBER: JP2002-327516
PRIOR FILING DATE: 2002-09-28
PRIOR APPLICATION NUMBER: JP2002-383869
PRIOR FILING DATE: 2002-12-09
NUMBER OF SEQ ID NOS: 27110
SEQ ID NO 17105
LENGTH: 228
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
FEATURE: OTHER INFORMATION: Z67364
FEATURE:
OTHER INFORMATION: Located on chromosome 11
FEATURE:
OTHER INFORMATION: Distance between a terminus base of telomere on


```

? OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
? OTHER INFORMATION: sequence : 13242364
? FEATURE:
? OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
? OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and
? OTHER INFORMATION: 5'-terminus of this base sequence : 225415
? OS-10-674-124A-17105

```

Query Match	88.2%	Score 19.4	DB 16	length 226
Best Local Similarity	95.2%	Pred. No. 30		
Matches	20	Conservative	0	Mismatches 1, Indels 0, Gaps 0
Qy	2	ATTTCGTTGGCTGCCTCCTG	22	
Db	191	ATTTCGTTGGCTGCCTCCTG	211	

RESULT 13
US-10-295-027-581
; Sequence 581, Application US/10295027
; Publication No. US20030232350A1
CURRENT INFORMATION

```

APPLICANT: Afar, Daniel
APPLICANT: Aziz, Natasha
APPLICANT: Gineberg, Wendy M.
APPLICANT: Glesh, Kurt C.
APPLICANT: Glyne, Richard
APPLICANT: Hevezl, Peter A.
APPLICANT: Mack, David H.
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan R.
APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
FILE REFERENCE: 018501-012500US
CURRENT FILING DATE: 2002-11-13
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/663,733
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: US 60/350,666
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US 60/335,394
PRIOR FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US 60/332,464
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: US 60/334,393
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: US 60/340,376
PRIOR FILING DATE: 2001-12-14
PRIOR APPLICATION NUMBER: US 60/347,211
PRIOR FILING DATE: 2002-01-08
PRIOR APPLICATION NUMBER: US 60/347,349
PRIOR FILING DATE: 2002-01-10
PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 581
LENGTH: 2190
TYPE: DNA
ORGANISM: Homo sapiens
US-10-295-027-581

```

	Query Match	88.2%	Score 19.4	DB 15,	Length 2199;
	Best Local Similarity	95.2%	Pred No. 39;		
	Matches	20; Conservative	0; Mismatches	1; Indels	0;
Qy	2	ATTTTGTTGGCTGCCTCCTG	22		
Db	2026	ATTTTGTTTCCTGCCTCCTG	2046		

```

RESULT 14
US-10-789-090-1
; Sequence 1, Application US/10789090
; Publication No. US20040223970A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Law, Debbie
; TITLE OF INVENTION: ANTIBODIES AGAINST SLC15A2 AND USES THEREOF
; FILE REFERENCE: 05882,0192,NPUS01
; CURRENT APPLICATION NUMBER: US/10/789,090
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2681
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-789-090-1

```

Query Match	Score	DB	Length
Best Local Similarity	95.2%	Pred. No. 35;	
Matches	20;	Conservative	0; Mismatches 1; Indels 0; Gaps 0;

Qy	2	ATTTCGTCCTGCTCCG	22
Db	2055	ATTTCGTCCTGCTCCG	2075

```

RESULT 15
US-09-964-824A-257
; Sequence 257 Application US/09964824A
; Patent No. US20020102531A1
; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signat
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964, 824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236, 033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236, 032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236, 028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 257
; LENGTH: 2685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-257

```

Query	Similarity	Score	DB	Length
Best Local	88.2%	19.4	9	2685
Matches	20	Conservative	0	Mismatches
			1	Indels
			0	Gaps

Search completed: February 9, 2005, 06:50:52
Job time : 31.6449 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 : Search time 262.041 Seconds
(Without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-3

Perfect score: 22
Sequence: 1 gattctgtctgctgctctctg 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hnc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gest1.*
9: gb_gest2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	100.0	303	7	W33932
2	22	100.0	430	7	W40649
3	22	100.0	536	7	W40649 mc43n12.r1
4	22	100.0	536	7	W40649 mc43n12.r1
5	22	100.0	536	7	W40649 mc43n12.r1
6	22	100.0	536	7	W40649 mc43n12.r1
7	22	100.0	536	7	W40649 mc43n12.r1
8	22	100.0	536	7	W40649 mc43n12.r1
9	22	100.0	536	7	W40649 mc43n12.r1
10	22	100.0	536	7	W40649 mc43n12.r1
11	22	100.0	536	7	W40649 mc43n12.r1
12	22	100.0	536	7	W40649 mc43n12.r1
13	22	100.0	536	7	W40649 mc43n12.r1
14	22	100.0	536	7	W40649 mc43n12.r1
15	22	100.0	536	7	W40649 mc43n12.r1
16	22	100.0	536	7	W40649 mc43n12.r1
17	22	100.0	536	7	W40649 mc43n12.r1
18	22	100.0	536	7	W40649 mc43n12.r1
19	22	100.0	536	7	W40649 mc43n12.r1
20	22	100.0	536	7	W40649 mc43n12.r1
21	22	100.0	536	7	W40649 mc43n12.r1
22	22	100.0	536	7	W40649 mc43n12.r1
23	22	100.0	536	7	W40649 mc43n12.r1
24	22	100.0	536	7	W40649 mc43n12.r1

25	20.4	92.7	449	7	R47024	R47024 Y350 Rat in
26	20.4	92.7	556	7	R46934	R46934 Y169 Rat in
27	20.4	92.7	570	9	AY419439	AY419439 Homo sapi
28	20.4	92.7	570	9	AY419440	AY419440 Pan trogl
29	20.4	92.7	1031	6	CB575508	CB575508 AGENCOURT
30	20.4	92.7	286	7	R46887	R46887 Y106 Rat in
31	19.4	88.2	293	7	R47143	R47143 Y79 Rat inc
32	19.4	88.2	298	7	R47143	R47143 Y79 Rat inc
33	19.4	88.2	303	7	R46903	R46903 Y124 Rat in
34	19.4	88.2	318	7	R47078	R47078 Y534 Rat in
35	19.4	88.2	355	7	R46940	R46940 Y174 Rat in
36	19.4	88.2	355	7	R46941	R46941 Y180 Rat in
37	19.4	88.2	373	7	R47135	R47135 Y722 Rat in
38	19.4	88.2	395	7	R47030	R47030 Y359 Rat in
39	19.4	88.2	458	2	AM295942	AM295942 UT-H-B12
40	19.4	88.2	461	1	AA054054	AA054054 Z148e05.r
41	19.4	88.2	467	7	R46947	R46947 Y186 Rat in
42	19.4	88.2	523	1	AA649247	AA649247 n835d11.8
43	19.4	88.2	621	2	AM974548	AM974548 EST386652
44	19.4	88.2	656	5	BQ017433	BQ017433 UT-H-DT1-
45	19.4	88.2	659	7	CF366828	CF366828 B41530 MA

ALIGNMENTS

RESULT 1
LOCUS W33932
DEFINITION mb54d02.r1 Soares mouse p3NMP19.5 Mus musculus cDNA clone
IMAGE:333219.5' similar to gb:M10095 (Mouse amelogenin (MOUSE)).
RNA sequence.
W33932
W33932.1 GI:1315837
EST.
VERSION W33932.1 GI:1315837
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

AUTHORS
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacey, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LIND, contact the
IMAGE Consortium (info@image.lind.gov) for further information.
WGI:214619
Seq primer: mob. REGA+ET
High quality sequence stop: 185.
Location/Qualifiers
1..303
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:333219"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_1lb="Soares mouse p3NMP19.5"
/note="Vector: pRT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTATACCAATCTGAGTGGAGCGCGCATTTTATTTTATTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI

FEATURES

source
FEATURES
source

ORIGIN

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Query Match 100.0%; Score 22; DB 7; Length 303;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTG 22
|||||
Db 81 GATTTGTTGCTGCTCTG 102

RESULT 2 430 bp mRNA linear EST 11-SEP-1996
W40649
LOCUS
DEFINITION
IMAGE:351335.5 similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.
ACCESSION
W40649.1 GI:1324982
VERSION
W40649
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:221335

Seq primer: ETPRimer
High quality sequence stop: 323.
Location/Qualifiers

FEATURES

source
1. 430
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:351335"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pRT3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match

100.0%; Score 22; DB 7; Length 430;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTG 22
|||||
Db 69 GATTTGTTGCTGCTCTG 90

RESULT 3 536 bp mRNA linear EST 11-SEP-1996
W36345
LOCUS
DEFINITION
IMAGE:334966.5 similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.
ACCESSION
W36345.1 GI:1318120
VERSION
W36345
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 536)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:216366

Seq primer: ETPRimer
High quality sequence stop: 359.
Location/Qualifiers

FEATURES

source
1. 536
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:334966"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pRT3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 100.0%; Score 22; DB 7; Length 536;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTG 22
|||||
Db 64 GATTTGTTGCTGCTCTG 85

RESULT 4

AY419441 585 bp DNA linear GSS 17-DEC-2003
LOCUS AY419441
DEFINITION Mus musculus AMELX gene, VIRUAL TRANSCRIPT, partial sequence.
ACCESSION AY419441
VERSION AY419441.1 GI:39775398
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 585)
Clar,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
AUTHORS 2 (bases 1 to 585)
Clar,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submision
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.
FEATURES
source
1..585
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>585
/gene="AMELX"
/locus_tag="HGM6891"
ORIGIN
Query Match 100.0%; Score 22; DB 9; length 585;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTG 22
Db 12 GATTTGTTGCTGCTCTG 33
RESULT 5
CD773419 645 bp mRNA linear EST 02-JUL-2003
LOCUS CD773419
DEFINITION AGENCOURT_14713373 NIH MGC_190 Mus musculus cDNA clone
ACCESSION CD773419
IMAGE:30501267 5', mRNA sequence.
VERSION CD773419.1 GI:32431921
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 645)
NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement: Yoshiko Yamada, Takashi Nakamura, NIDCR
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDCM223 row: k column: 04
High quality sequence stop: 611.
FEATURES
source
1..645
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30501267"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_190"
/note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site_1: SfiI (ggccgctggcc); Site_2: SfiI (ggccgctggcc); Non-normalized full-length enriched library 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.71 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) Corp."
ORIGIN
Query Match 100.0%; Score 22; DB 6; length 645;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTG 22
Db 84 GATTTGTTGCTGCTCTG 105
RESULT 6
CB056709 707 bp mRNA linear EST 17-JAN-2003
LOCUS CB056709
DEFINITION NISC_J11905.W1 Scores NMBP13-15 Mus musculus cDNA clone
ACCESSION CB056709
IMAGE:4848584 5', mRNA sequence.
VERSION CB056709.1 GI:27794996
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
CDNA Library Preparation: The I.M.A.G.E. Consortium/LNL
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
info@image.llnl.gov
Plate: LHAM10779 row: N column: 9
Seg primer: T7 primer.
FEATURES
source
1..707
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4848584"
/issue_type="Pituitary gland"
/dev_stage="juvenile, 13-15 days"
/lab_host="DH10B (phage-resistant)"

/clone.lib="Soares NMBP13-15"
 /note="Organ: brain; Vector: pT73D-Pact; Site 1: NotI;
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-
 01190 (GT) primer
 5'-AAGTGAAGAAATTCGCGCGCCGCTGACCAAGTATTTTATTTT-3';
 double-stranded cDNA was ligated to EcoRI adaptors
 5'-ATTGCGACGAG-3' and 5'-CCTGCGCG-3' (Pharmacia),
 digested with NotI and cloned into the NotI and EcoRI
 sites of the pT73D-Pact vector. Library went through one
 round of normalization, and was constructed in the
 laboratory of M. Bento Soares (University of Iowa)."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 707;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTTGTTGCTGCTGCTGCTG 22
 58 GATTTGTTGCTGCTGCTGCTG 79

Db

RESULT 7
 AK029358 812 bp mRNA linear HTC 03-APR-2004
 LOCUS AK029358
 DEFINITION Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
 library, clone:4833404E21 product:amelogenin, full insert sequence.
 ACCESSION AK029358
 VERSION AK029358.1 GI:26325321
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishise, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)

REFERENCE
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 812)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE

Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers

FEATURES

source

1..812
 /organism="Mus musculus"
 /mol_type="RNA"
 /strain="CS7BL/6J"
 /db_xref="FANTOM:DB:4833404E21"
 /db_xref="taxon:10090"
 /clone="4833404E21"
 /issue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 73..663
 /note="unnamed protein product; amelogenin (MGD|MG1:88005,
 GB|D11768, evidence: BLASTN, 99%, match=799)
 putative"
 /codon_start=1
 /protein_id="AAC26415.1"
 /db_xref="GI:26325322"
 /translation="MTGTLPRACLLGAFAFPLPPHSGSPYINLSYEVLTPLKMYQS
 MIRQYPSYGYEPMGWLHQIIPVLSQHPSPHLDQHNHLPVPAQPPVAPQPPM
 PVPGRHSMTPTQHQPNIPPSAQCPFPQPPQIPQSHQPMQFOSPLHMOPLAQ
 PPLPPLFPMQPLSLPLPLPLEAMPATDKTGREVD"
 793..798
 /note="putative"
 812
 /note="putative"
 /note="putative"

CDS

polyA_signal

polyA_site

ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 812;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTTGTTGCTGCTGCTGCTG 22
 84 GATTTGTTGCTGCTGCTGCTG 105

Db

RESULT 8
 CB588212 843 bp mRNA linear EST 03-APR-2003
 LOCUS CB588212
 DEFINITION AGENCOURT 12771359 NIH MGC 136 Mus musculus cDNA clone
 IMAGE:30293239 5', mRNA sequence.
 ACCESSION CB588212
 VERSION CB588212.1 GI:29506068
 KEYWORDS EST.

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 843)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: NDAM318 row: 9 column: 08
High quality sequence stop: 636.

FEATURES
Location/Qualifiers
1..843
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293239"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5', 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 843;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 GATTTGTTGGCTGCTCCTG 22
|||||
51 GATTTGTTGGCTGCTCCTG 72
|||||

Db
51 GATTTGTTGGCTGCTCCTG 72
|||||

RESULT 9
CB588525 859 bp mRNA linear EST 03-APR-2003
AGENCOURT 12567944 NIH MGC 136 Mus musculus CDNA clone
IMAGE:30289751 5', mRNA sequence.

ACCESSION
CB588525
CB588525.1 GI:29506381
EST.

KEYWORDS
Mus musculus (house mouse)

SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 859)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>
Plate: NDAM329 row: e column: 24
High quality sequence stop: 763.

FEATURES
Location/Qualifiers
1..859
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5', 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 859;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 GATTTGTTGGCTGCTCCTG 22
|||||
70 GATTTGTTGGCTGCTCCTG 91
|||||

Db
70 GATTTGTTGGCTGCTCCTG 91
|||||

RESULT 10
CB589177 865 bp mRNA linear EST 03-APR-2003
AGENCOURT 12770446 NIH MGC 136 Mus musculus CDNA clone
IMAGE:30290534 5', mRNA sequence.

ACCESSION
CB589177
CB589177.1 GI:29507033
EST.

KEYWORDS
Mus musculus (house mouse)

SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 865)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: NDAM331 row: f column: 15
High quality sequence stop: 630.

FEATURES
Location/Qualifiers
1..865
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290534"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5', 4 and 1 limb and jaw

ORIGIN

equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC library."

Query Match 100.0%; Score 22; DB 6; Length 865;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTG 22
Db 69 GATTTGTTGCTGCTGCTG 90

RESULT 11
LOCUS CB587051 886 bp mRNA linear EST 03-APR-2003
DEFINITION AGENCOURT 12807074 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30295123 5', mRNA sequence.
ACCESSION CB587051
VERSION CB587051.1 GI:29504907
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 886)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDAM343 row: e column: 20
High quality sequence stop: 619.
Location/Qualifiers

FEATURES

source

1..886
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30295123"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 886;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTG 22
Db 95 GATTTGTTGCTGCTGCTG 116

RESULT 12
LOCUS CB589251 889 bp mRNA linear EST 03-APR-2003
DEFINITION AGENCOURT 12780700 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30287069 5', mRNA sequence.

Query Match 100.0%; Score 22; DB 6; Length 889;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTG 22
Db 89 GATTTGTTGCTGCTGCTG 110

RESULT 13
LOCUS CB574837 891 bp mRNA linear EST 02-APR-2003
DEFINITION AGENCOURT 12973279 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30294719 5', mRNA sequence.
ACCESSION CB574837
VERSION CB574837.1 GI:29494367
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 891)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/

FEATURES

source

1..889
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30287069"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 889;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTG 22
Db 89 GATTTGTTGCTGCTGCTG 110

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDAM342 row: d column: 24
High quality sequence start: 19
High quality sequence stop: 658.
Location/Qualifiers

FEATURES

source

1. .891
/organism="Mus musculus"
/mol_type="RNA"
/db_xref="taxon:10090"
/clone="IMAGE:30294719"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 891;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22
|||||
DB 90 GATTTGTTGCTGCTCCTG 111

RESULT 14
CB590451 913 bp mRNA linear EST 03-APR-2003
LOCUS
DEFINITION
AGENCOURT 12770290 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30286944 5', mRNA sequence.
CB590451
VERSION
CB590451.1 GI:29508307
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 913)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDAM322 row: a column: 01
High quality sequence start: 12
High quality sequence stop: 640.
Location/Qualifiers

FEATURES

source

1. .919
/organism="Mus musculus"
/mol_type="RNA"
/db_xref="taxon:10090"
/clone="IMAGE:30286944"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size

source

1. .913
/organism="Mus musculus"
/mol_type="RNA"
/db_xref="taxon:10090"
/clone="IMAGE:30286944"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 913;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCCTG 22
|||||
DB 82 GATTTGTTGCTGCTCCTG 103

RESULT 15
CB587332 919 bp mRNA linear EST 03-APR-2003
LOCUS
DEFINITION
AGENCOURT 12971606 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30291971 5', mRNA sequence.
CB587332
VERSION
CB587332.1 GI:29505188
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 919)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDAM335 row: b column: 12
High quality sequence start: 216
High quality sequence stop: 325.
Location/Qualifiers

FEATURES

source

1. .919
/organism="Mus musculus"
/mol_type="RNA"
/db_xref="taxon:10090"
/clone="IMAGE:30291971"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size

1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 919;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTTGTTGGCTGCTGCTG 22
|||
Db 69 GATTTGTTGGCTGCTGCTG 90

Search completed: February 8, 2005, 22:17:57
Job time : 267.041 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 122.665 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22

Sequence: 1 gctctgtctactgtcttaacc 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : GenEmbl.*

1: gb_ba.*
2: gb_ntg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_dl.*
9: gb_pr.*
10: gb_ro.*
11: gb_str.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18.8	85.5	6451	4 AB091789	AB091789 Bos tauru
2	18.8	85.5	208214	10 AL645908	AL645908 Mouse DNA
3	18.4	83.6	184836	2 AC140360	AC140360 Mus muscu
4	18.4	81.8	162927	10 AC115302	AC115302 Mus muscu
5	18.8	81.8	209617	10 AC139300	AC139300 Mus muscu
6	17.8	80.9	108799	9 AC090989	AC090989 Homo sapi
7	17.8	80.9	168637	5 AL844187	AL844187 zebrafish
8	17.8	80.9	176148	9 AC010798	AC010798 Homo sapi
9	17.8	80.9	177151	2 AC132312	AC132312 Mus muscu
10	17.8	80.9	178139	2 AC118710	AC118710 Mus muscu
11	17.8	80.9	178825	2 AP001203	AP001203 Homo sapi
12	17.8	80.9	188972	10 AC101658	AC101658 Homo sapi
13	17.8	80.9	208244	2 BX470182	BX470182 Danio rer
14	17.8	80.9	213071	2 AC110680	AC110680 Rattus no
15	17.8	80.9	223538	2 AC087150	AC087150 Mus muscu
16	17.8	80.9	238660	2 AC132354	AC132354 Mus muscu
17	17.8	80.9	244868	2 AC095331	AC095331 Rattus no
18	17.8	80.9	245528	2 AC133363	AC133363 Rattus no
19	17.8	80.9	251219	2 CR388422	CR388422 Danio rer

20	17.8	80.9	270721	2 AC105775	AC105775 Rattus no
21	17.4	79.1	71701	2 AF254984	AF254984 Homo sapi
22	17.4	79.1	75212	2 AC136287	AC136287 Medicago
23	17.4	79.1	86554	2 AC090280	AC090280 Homo sapi
24	17.4	79.1	105191	9 AL161655	AL161655 Human DNA
25	17.4	79.1	122924	2 AF207954	AF207954 Homo sapi
26	17.4	79.1	124630	9 AC090731	AC090731 Homo sapi
27	17.4	79.1	146734	10 AL590415	AL590415 Mouse DNA
28	17.4	79.1	167232	5 AL929379	AL929379 zebrafish
29	17.4	79.1	172875	9 AC084301	AC084301 Homo sapi
30	17.4	79.1	195119	5 BX649594	BX649594 zebrafish
31	17.4	79.1	214811	10 AL662895	AL662895 Mouse DNA
32	17.4	79.1	215615	2 AC098139	AC098139 Rattus no
33	17.4	79.1	326786	2 AC022365	AC022365 Homo sapi
34	17.2	78.2	334	8 AY142645	AY142645 Arabidops
35	17.2	78.2	363	6 BD248010	BD248010 5' ESTs f
36	17.2	78.2	363	6 CO771973	CO771973 Sequence
37	17.2	78.2	397	11 HS2822E1	Z66933 H. sapiens D
38	17.2	78.2	479	6 AR415985	AR415985 Sequence
39	17.2	78.2	479	6 AX972819	AX972819 Sequence
40	17.2	78.2	479	6 BD111538	BD111538 EST and e
41	17.2	78.2	488	8 AY034985	AY034985 Arabidops
42	17.2	78.2	640	8 AY086566	AY086566 Arabidops
43	17.2	78.2	1165	9 BC000735	BC000735 Homo sapi
44	17.2	78.2	1645	9 HSY13613	YS13613 Homo sapien
45	17.2	78.2	1935	9 D83730	D83730 Homo sapien

ALIGNMENTS

RESULT 1
AB091789/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus (cow)
Bos taurus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Ruminantia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 6451)
Iwase,M., Satte,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies(Sokendai), Department of Biosystems Science,
Shonan Kikusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@min@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
Location/Qualifiers
1. 6451
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/sex="male"
238. 6451
/gene="AMELX"
join(238..340,1650..1661)
/gene="AMELX"
join(1662..1715,4173..4220,5560..5601,5665..5709,
5987..6451)
/gene="AMELX"
/codon_start=1
/product="amelogenin"

/protein_id="BAC66108.1"
 /db_xref="gi:29126031"
 /translation="MGTRITLACILGAAFPMLPDPHGGYINFSYENSYPQISID
 ETALVLPKMTQSMIRHYPISYIEPMGWLHHQIIPVSOQTPQNHAIQPHHILPM
 VPAQPPVPOQPMKVPQSHMTPTQHHQPMPLPAQOPFOPOSIQPPHQPQPPH
 LQPMQOPLOPLOPQPPVHPIQPLPQPPPIPIFPMQPLPPLPDLPLAMWPT
 "

ORIGIN
 Query Match 85.5%; Score 18.8; DB 4; Length 6451;
 Best Local Similarity 90.9%; Pred. No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GATCTGTACTGTCTTAAACC 22
 |||||
 Db 1821 GATCATGTACTGTCTTAAACC 1800

RESULT 2
 AL645908 208214 bp. DNA linear ROD 05-APR-2002
 LOCUS Mouse DNA sequence from clone RP23-24A24 on chromosome 11, complete
 DEFINITION
 AL645908
 VERSION AL645908.9 GI:20068640
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Pearce, A.
 TITLE Direct Submission
 JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 7, 2002 this sequence version replaced gi:17976640.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 from the RPT-23 Mouse PAC library
 constructed by the group of Pletier de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.

FEATURES

source
 1..208214
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-24A24"
 /clone_1fb="RPT-23"

ORIGIN

Query Match 85.5%; Score 18.8; DB 10; Length 208214;
 Best Local Similarity 90.9%; Pred. No. 66;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGTACTGTCTTAAACC 22
 |||||
 Db 80649 GATCTGTACTGTCTTAAACC 80670

RESULT 3
 AC140360/c 184836 bp DNA linear HTG 12-JUN-2004
 LOCUS Mus musculus chromosome 12 clone RP24-24IG4, WORKING DRAFT
 DEFINITION
 AC140360
 SEQUENCE. 5 unordered pieces.
 AC140360
 VERSION AC140360.2 GI:48675488
 KEYWORDS HTG, HTGS, PHASIT; HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Wilson, R.K.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 184836)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 184836)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUN-2004) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 On Jun 12, 2004 this sequence version replaced gi:28475618.

COMMENT

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@wustl.edu
 Project Information
 Center project name: W_BB0241G04
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 182119 bases at least Q40
 Consensus quality: 182305 bases at least Q30
 Consensus quality: 182480 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4967: contig of 4967 bp in length
 * 4968 5067: gap of unknown length
 * 5068 12306: contig of 7239 bp in length
 * 12307 12406: gap of unknown length
 * 12407 35592: contig of 23186 bp in length
 * 35593 35692: gap of unknown length
 * 35693 88170: contig of 52478 bp in length
 * 88171 88271: gap of unknown length
 * 88271 184836: contig of 96566 bp in length.

FEATURES

source
 1..184836
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"

```

misc_feature      /chromosome="12"
                  /clone="RP24-241G4"
                  1..4967
misc_feature      /note="assembly_name:Contig951"
                  5068..12306
misc_feature      /note="assembly_name:Contig952"
                  12407..35592
misc_feature      /note="assembly_name:Contig953"
                  35693..88170
misc_feature      /note="assembly_name:Contig954"
                  88271..184836
misc_feature      /note="assembly_name:Contig955"

```

ORIGIN

```

Query Match      83.6%; Score 18.4; DB 2; Length 184836;
Best Local Similarity 95.0%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

3 TCTTGACTGCTTAAACC 22
| ||||| ||||| |||||
Db 408 TCTTGACTGCTTAAACC 389

```

```

RESULT 4
AC115302      162927 bp  DNA  linear  ROD 27-NOV-2003
LOCUS      Mus musculus BAC clone RP24-107D24 from chromosome 18, complete
SEQUENCE.
AC115302      GI:30142728
HTG.
Mus musculus (house mouse)
Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 162927)
Boatright, E., Walligorski, J., Haakenson, W. and Doeber, A.
The sequence of Mus musculus BAC clone RP24-107D24
Unpublished (2001)
2 (bases 1 to 162927)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 162927)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (16-MAR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 162927)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 162927)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (27-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 162927)
Wilson, R.
Direct Submission
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 27, 2003 this sequence version replaced gi:22475848.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.wustl.edu
-----
Summary Statistics
-----
Center project name: M_BB0107D24
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC131796.

FEATURES

source

```

1..162927
Location/Qualifiers

```

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"

```

```

/map="18"
/clone="RP24-107D24"
/clone_1lb="RPCI-24"

```

```

1..1209
/rpc_family="L1"

```

```

1..1136
/rpc_family="L1"

```

```

/note="Sequence derived from PCR product of project DNA."
repeat_region
1209..1556
/rpc_family="L1"

```

```

1617..1938
/rpc_family="MaLR"

```

```

1941..2463
/rpc_family="MaLR"

```

```

2506..3065
/rpc_family="MaLR"

```

```

3070..3209
/rpc_family="MaLR"

```

```

3504..3754
/rpc_family="Alu"

```

```

3755..4362
/rpc_family="MaLR"

```

```

4363..5160
/rpc_family="ERV1"

```

```

5164..5315
/rpc_family="ERV1"

```

```

8477..9064
/rpc_family="ERV1"

```

```

9067..9644
/rpc_family="MaLR"

```

```

9645..9970
/rpc_family="MaLR"

```

```

10028..10128
/rpc_family="MaLR"

```

```

10128..10512
/rpc_family="BMER16"

```

```

10642..10721
/rpc_family="BMER16"

```

```

10870..11144
/rpc_family="MER2_type"

```

```

repeat_region
/rpc_family="B4"

```

```

repeat_region 11752..14937
                /rpt_family="L1"
repeat_region 14938..15072
                /rpt_family="ERV1"
repeat_region 18731..19077
                /rpt_family="MALR"
repeat_region 19931..20167
                /rpt_family="L1"
repeat_region 20151..20262
                /rpt_family="L1"
repeat_region 20309..20620
                /rpt_family="L1"
repeat_region 20727..22087
                /rpt_family="L1"
repeat_region 22105..22300
                /rpt_family="L1"
repeat_region 22348..22642
                /rpt_family="RMR19A"
repeat_region 22846..22991
                /rpt_family="L1"
repeat_region 23027..23427
                /rpt_family="L1"
repeat_region 23410..23734
                /rpt_family="L1"
repeat_region 26778..27659
                /rpt_family="L1"
repeat_region 28064..28180
                /rpt_family="RMR13B"
repeat_region 28179..28783
                /rpt_family="RMR13B"
repeat_region 28558..28794
                /rpt_family="ERVK"
repeat_region 29543..29690
                /rpt_family="B4"
repeat_region 29723..30108
                /rpt_family="MALR"
repeat_region 30799..30977
                /rpt_family="ERV1"
repeat_region 32579..32823
                /rpt_family="L1"
repeat_region 33660..33807
                /rpt_family="Alu"
repeat_region 34460..34606
                /rpt_family="Alu"
repeat_region 34639..34846
                /rpt_family="B2"
repeat_region 35192..35438
                /rpt_family="MALR"
repeat_region 35703..35938
                /rpt_family="MALR"
repeat_region 35993..36399
                /rpt_family="MALR"
repeat_region 37273..37677
                /rpt_family="MALR"
repeat_region 38289..38533
                /rpt_family="MALR"
repeat_region 38542..39278
                /rpt_family="L1"
repeat_region 39297..40291
                /rpt_family="L1"
repeat_region 40292..40668
                /rpt_family="MALR"
repeat_region 40669..40815
                /rpt_family="Alu"
repeat_region 40816..40896
                /rpt_family="MALR"
repeat_region 40897..41065
                /rpt_family="L1"
repeat_region 41715..41994
                /rpt_family="L1"
repeat_region 42350..42542
                /rpt_family="B2"
repeat_region 43595..43724

```

```

repeat_region 43981..44067
                /rpt_family="Alu"
repeat_region 43990..44512
                /rpt_family="Alu"
repeat_region 45020..45260
                /rpt_family="MALR"

Query Match      81.8%; Score 18; DB 10; Length 162927;
Best Local Similarity 100.0%; Pred.No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      2  ATCTGTACTGCTTAA 19
Db      134116 ATCTGTACTGCTTAA 134133

RESULT 5
AC139300/c
LOCUS      209617 bp    DNA    linear    ROD 25-NOV-2003
DEFINITION Mus musculus BAC clone RP23-258L8 from chromosome 18, complete
sequence.
AC139300
VERSION    AC139300.4  GI:37361067
KEYWORDS   HTG.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 209617)
REFERENCE  1
  AUTHORS  Levy, A.
  TITLE    The sequence of Mus musculus BAC clone RP23-258L8
  JOURNAL  Unpublished (2001)
REFERENCE  2
  AUTHORS  Wilson, R.
  TITLE    Sequencing of Mus musculus
  JOURNAL  Unpublished (2001)
REFERENCE  3
  AUTHORS  McPherson, J.D. and Waterston, R.H.
  TITLE    Direct Submission
  JOURNAL  Submitted (29-JAN-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            4 (bases 1 to 209617)
REFERENCE  4
  AUTHORS  Wilson, R.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            5 (bases 1 to 209617)
REFERENCE  5
  AUTHORS  Wilson, R.K.
  TITLE    Direct Submission
  JOURNAL  Submitted (02-OCT-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            6 (bases 1 to 209617)
REFERENCE  6
  AUTHORS  Wilson, R.
  TITLE    Direct Submission
  JOURNAL  Submitted (25-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 2, 2003 this sequence version replaced gi:30911162.

COMMENT
  ----- Genome Center
  Center: Washington University Genome Sequencing Center
  Center code: WUGSC
  Web site: http://genome.wustl.edu
  Contact: submissions@wustl.wustl.edu
  ----- Summary Statistics
  -----
  Center project name: M_BA0258L08

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate


```

repeat_region      /rpt_family="MADE1"
                    17547..17567
repeat_region      /rpt_family="AT_rich"
                    complement(18293..18508)
repeat_region      /rpt_family="MLTJU"
                    18941..19018
repeat_region      /rpt_family="MIR"
                    19077..19098
repeat_region      /rpt_family="AT_rich"
                    19453..19756
repeat_region      /rpt_family="AluJo"
                    20661..20872
repeat_region      /rpt_family="L2"
                    complement(23626..23836)
repeat_region      /rpt_family="MIR"
                    25418..25458
repeat_region      /rpt_family="AT_rich"
                    27001..27112
repeat_region      /rpt_family="LIME"
                    complement(27235..27408)
repeat_region      /rpt_family="MERSB"
                    27508..27533
repeat_region      /rpt_family="AT_rich"
                    29143..29441
repeat_region      /rpt_family="AluSc"
                    complement(30072..30333)
repeat_region      /rpt_family="LMC/D"
                    complement(30971..31057)
repeat_region      /rpt_family="MLTK"
                    31882..31946
repeat_region      /rpt_family="(CAA)n"
                    complement(32318..32545)
repeat_region      /rpt_family="MER46A"
                    33184..33214
repeat_region      /rpt_family="AT_rich"
                    33562..33613
repeat_region      /rpt_family="L2"
                    33682..33816
repeat_region      /rpt_family="(TA)n"
                    complement(33817..33938)
repeat_region      /rpt_family="FLAM_C"
                    34204..34421
repeat_region      /rpt_family="MIR"
                    34576..34607
repeat_region      /rpt_family="AT_rich"
                    complement(34716..34944)
repeat_region      /rpt_family="MIR"
                    35815..35875
repeat_region      /rpt_family="(TA)n"
                    complement(35975..35979)
unrepeated          /note="30 qual SNGI region"
repeat_region      36347..36377
repeat_region      /rpt_family="AT_rich"
                    36547..36611
repeat_region      /rpt_family="AT_rich"
                    37724..37744
repeat_region      /rpt_family="AT_rich"
                    38576..38603
repeat_region      /rpt_family="AT_rich"
                    40354..40383
repeat_region      /rpt_family="AT_rich"
                    41112..41139
repeat_region      /rpt_family="AT_rich"
                    complement(43246..46180)
repeat_region      /rpt_family="L2"

```

```

Query Match      80.9%; Score 17.8; DB 9; Length 108799;
Best Local Similarity 90.5%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2 ATCTGTACTGTCTTAACCC 22
DB      84043 ATCTGTACTGTCTTAACCC 84023

```

```

RESULT 7
AL844187          168637 bp      DNA      linear      VRT 13-DEC-2003
LOCUS            Zebrafish DNA sequence from clone DKEY-221D18 in linkage group 16,
DEFINITION       complete sequence.
ACCESSION        AL844187.13 GI:39919157
VERSION          HTG.
KEYWORDS         Danio rerio (zebrafish)
SOURCE           Danio rerio
ORGANISM         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                  Cypriniformes; Cyprinidae; Danio.
REFERENCE        Submitted (13-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
AUTHORS          Cambridgehire, CB10 1SA, UK. E-mail enquiries:
TITLE            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
JOURNAL          On Dec 15, 2003 this sequence version replaced gi:29119912.
COMMENT          ----- Genome Center
                  Center: Wellcome Trust Sanger Institute
                  Center code: SC
                  Web site: http://www.sanger.ac.uk
                  Contact: zfish-help@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissions corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep DKEY-221D18 is from a Zebrafish BAC library

VECTOR: pindigBAC-5

Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml.

```

FEATURES
    source
        1..168637
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-221D18"
            /clone_1ib="Daniokey"

```

ORIGIN

```

Query Match      80.9%; Score 17.8; DB 5; Length 168637;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      2 ATCTGTACTGTCTTAACCC 22

```

Db 106375 ATCTGTACTGTTTACACCC 106395

RESULT 8
AC010798/c 176148 bp DNA linear PRI 30-MAY-2002

LOCUS Homo sapiens, clone RP11-470B24, complete sequence.
DEFINITION AC010798
AC010798
AC010798.9 GI:19683107
KEYWORDS HTG
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 176148)
Birtten, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-470B24
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 176148)
Birtten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Brown, A., Castle, A., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Cooke, P., Dekrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galgan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Margulis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teefaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
TITLE Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 176148)
Birtten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campiolano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Gard, S., Gerd, S., Goyette, W., Gage, D., Galgan, J., Gardyna, S.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamazares, R.,
Landers, T., Lehoczky, J., Levine, R., Lindblad-Ton, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, P., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Teefaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
TITLE Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 4 (bases 1 to 176148)
Birtten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campiolano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Galgan, J., Gardyna, S., Ginde, S., Gerd, S., Goyette, M., Gage, D.,
Galgan, J., Gardyna, S., Ginde, S., Gerd, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacombe, K.,
Lamazares, R., Landers, T., Lehoczky, J., Levine, R., Lindblad-Ton, K.,

Li, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
Schnupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Teefaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
TITLE Submitted (30-MAY-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 22, 2002 this sequence version replaced gi:9280739.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1027
Center clone name: 470_B_24

FEATURES
source

Location/Qualifiers
1. 176148
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-470B24"
/clone_lib="RP11-470B24 Human Male BAC"
complement(222..374)
/rpt_family="HMLb"
400..479
/rpt_family="AT-rich"
complement(629..792)
/rpt_family="HMLb"
complement(915..1074)
/rpt_family="ALu0"
1075..1123
/rpt_family="TTTA)n"
complement(1124..1238)
/rpt_family="ALu0"
complement(3626..3918)
/rpt_family="HMLb"
complement(4449..4895)
/rpt_family="HMLb"
complement(5090..5312)
/rpt_family="MER58A"
complement(5566..5858)
/rpt_family="ALu0b"
complement(8060..8226)
/rpt_family="L2"
8523..8580
/rpt_family="L2"
complement(8581..8991)
/rpt_family="MSTB"
8992..9028
/rpt_family="L2"
complement(9914..10109)
/rpt_family="L2"
complement(12003..12305)
/rpt_family="ALu0x"
complement(12340..12857)
/rpt_family="L1MC4a"
complement(12951..13026)
/rpt_family="L1MC4a"
complement(13027..13174)
/rpt_family="FLAM_C"

Query Match	80.9%	Score 17.8;	DB 9;	Length 176148;
Beat Local Similarity	90.5%	Pred. No. 2.2e+02;		
Matches 19;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
<p>QY 2 ATCTGTACTGCTTAAACC 22</p> <p> </p> <p>Db 11479 ATCTTACTGCTTAAACC 11459</p>				
repeat_region	/rpt_family="L13175"	complement(131175)		
repeat_region	/rpt_family="L1MC4a"	complement(13571)		
repeat_region	/rpt_family="L1MD3"	complement(14537)		
repeat_region	/rpt_family="L2"	complement(14894)		
repeat_region	/rpt_family="MIR"	complement(15083)		
repeat_region	/rpt_family="L2"	complement(15861)		
repeat_region	/rpt_family="MLT1B"	complement(18466)		
repeat_region	/rpt_family="AT_rich"	complement(20256)		
repeat_region	/rpt_family="L2"	complement(20479)		
repeat_region	/rpt_family="L1MED"	complement(20589)		
repeat_region	/rpt_family="L1MED"	complement(20733)		
repeat_region	/rpt_family="L1MED"	complement(20858)		
repeat_region	/rpt_family="FLAM_C"	complement(20859)		
repeat_region	/rpt_family="L1MED"	complement(20998)		
repeat_region	/rpt_family="L2"	complement(21048)		
repeat_region	/rpt_family="L2"	complement(21422)		
repeat_region	/rpt_family="MLT1A1"	complement(21423)		
repeat_region	/rpt_family="L2"	complement(21803)		
repeat_region	/rpt_family="L2"	complement(22499)		
repeat_region	/rpt_family="L1M4B"	complement(22502)		
repeat_region	/rpt_family="L2"	complement(22656)		
repeat_region	/rpt_family="L2"	complement(22871)		
repeat_region	/rpt_family="AT_rich"	complement(22910)		
repeat_region	/rpt_family="L2"	complement(22962)		
repeat_region	/rpt_family="L2"	complement(24214)		
repeat_region	/rpt_family="AT_rich"	complement(24232)		
repeat_region	/rpt_family="L1M4"	complement(25908)		
repeat_region	/rpt_family="L1M4"	complement(25974)		
repeat_region	/rpt_family="L1M4"	complement(28106)		
repeat_region	/rpt_family="AT_rich"	complement(28134)		
repeat_region	/rpt_family="AT_rich"	complement(28528)		
repeat_region	/rpt_family="AT_rich"	complement(28757)		
repeat_region	/rpt_family="L1M3A"	complement(29361)		
repeat_region	/rpt_family="L1M3A"	complement(30915)		
repeat_region	/rpt_family="L1M3A"	complement(31357)		
repeat_region	/rpt_family="L1M3A"	complement(31489)		
repeat_region	/rpt_family="L1M3A"	complement(31567)		
repeat_region	/rpt_family="MIR20"	complement(31841)		

```

ACCESSION      SEQUENCE, 3 unordered pieces.
VERSION        AC132312
KEYWORDS       HTG, HTGS_PHASE1, HTGS_DRAFT, HTGS_ACTIVEFIN.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia, Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE          1 (bases 1 to 177151)
JOURNAL        The sequence of Mus musculus clone
AUTHORS        Wilson,R.K.
TITLE          Unpublished
JOURNAL        2 (bases 1 to 177151)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          Direct Submission
JOURNAL        Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS        Parkway, St. Louis, MO 63108, USA
TITLE          3 (bases 1 to 177151)
JOURNAL        Wilson,R.K.
REFERENCE      Direct Submission
AUTHORS        Submitted (18-JUN-2004) Genome Sequencing Center, 4444 Forest Park
TITLE          Parkway, St. Louis, MO 63108, USA
JOURNAL        On Jun 18, 2004 this sequence version replaced gi:37806543.

COMMENT        -----
                Genome Center
                -----
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
                -----
                Project Information
                -----
                Center project name: M_BB0298A20
                -----
                Summary Statistics
                -----
                Sequencing vector: M13, 0%
                Sequencing vector: Plasmid: 100%
                Chemistry: Dye-primer ET: 0% of reads
                Chemistry: Dye-terminator Big Dye, 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 174821 bases at least Q40
                Consensus quality: 175115 bases at least Q30
                Consensus quality: 175332 bases at least Q20
                -----
                * NOTE: This is a 'working draft' sequence. It currently
                * consists of 3 contigs. The true order of the pieces
                * is not known and their order in this sequence record is
                * arbitrary. Gaps between the contigs are represented as
                * runs of N, but the exact sizes of the gaps are unknown.
                * This record will be updated with the finished sequence
                * as soon as it is available and the accession number will
                * be preserved.
                -----
                * 1 18378: contig of 18378 bp in length
                * 18379 18478: gap of unknown length
                * 18479 70874: contig of 52396 bp in length
                * 70875 70974: gap of unknown length
                * 70975 177151: contig of 106177 bp in length.
                Location/Qualifiers
                1. 177151
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /chromosome="12"
                /clone="RP24-298A20"
                1. 18378
                /note="assembly_name:Contig2"
                18479. 70874
                /note="assembly_name:Contig3
                clone_end:17
                vector_side:right"
                70975. 177151
                /note="assembly_name:Contig4
                clone_end:SP6
                vector_side:right"

```

Query Match 80.9%; Score 17.8; DB 2; Length 177151;
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGTACTGTCTTAACC 21
 |||||
 Db 111099 GGTCTGTACTGTCTTAAGCC 111119

RESULT 10
 AC118710 178139 bp DNA linear HTG 29-APR-2004
 Mus musculus chromosome 15 clone RP24-121C9 map 15, *** SEQUENCING
 IN PROGRESS *** 2 ordered pieces.

AC118710
 AC118710.6 GI:46849683
 HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 178139)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus chromosome 15, clone RP24-121C9
 Unpublished
 2 (bases 1 to 178139)
 Anderson, S., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Camarata, J., Campobiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kellis, C., Labocque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Teste, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vasilev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 178139)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Anderson, S., Arachchi, H., Batma, N., Baetien, V.,
 Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J.,
 Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B.,
 DeArrellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L.,
 Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D.,
 Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hages, B.,
 Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T.,
 Levine, R., Lindblad-Toh, K., Liu, G., Liu, X., Liu, A., Mabbitt, R.,
 Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Menues, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nguyen, T., Nicol, R., Norbu, C.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C.,
 Retta, R., Riese, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Stubbs, M., Talamas, J., Teste, S., Theodore, J.,
 Topham, K., Travers, M., Vasilev, H., Venkataraman, V.S., Viel, R.,
 Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L.,

TITLE JOURNAL COMMENT

Zimmer, A. and Zody, M.

Direct Submission

Submitted (29-APR-2004) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 29, 2004 this sequence version replaced gi:46367856.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@road.mit.edu

Project Information

Center project name: 121_C_9

Center clone name: 121_C_9

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 66459: contig of 66459 bp in length

* 66460 66559: gap of 100 bp

* 66560 178139: contig of 111580 bp in length.

Location/Qualifiers

1. 178139

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP24-121C9"

/clone_lib="RPCT-24 Male Mouse BAC"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 178139;
 Best Local Similarity 90.5%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGTACTGTCTTAACC 22
 |||||
 Db 93272 ATCTGTACTGTCTTAAGCC 93292

RESULT 11
 AP001203 178825 bp DNA linear HTG 29-JUL-2000
 LOCUS
 DEFINITION
 Homo sapiens chromosome 18 clone RP11-845G3 map 18q12, WORKING
 DRAFT SEQUENCE, 6 unordered pieces.
 AP001203
 AP001203.3 GI:9588070
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 178825)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 178,825 genomic DNA of 18q12
 Published Only in Database (2000)
 2 (bases 1 to 178825)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (18-FEB-2000) Masahito Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-15-1 Kitaato, Sagamihara, Kanagawa 228-8555, Japan

COMMENT

(E-mail: hactori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-42-778-9923, Fax: 81-42-778-9924)
On Jul 31, 2000 this sequence version replaced gi:8117649.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hactori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraT18

Center clone name: RP11-845G3

----- Summary Statistics

Sequencing vector: PCR products: 100% of reads

Chemistry: Dye-terminator RT-amersham: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 177242 bases at least Q40

Consensus quality: 177817 bases at least Q30

Consensus quality: 178089 bases at least Q20

Insert size: 178325; sum-of-coverage

Quality coverage: 11.49x in Q20 bases; sum-of-coverage

----- NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```
1 79432 contig of 79432 bp in length
79533 120070 contig of 40538 bp in length
120171 152785 contig of 32615 bp in length
152886 164489 contig of 11604 bp in length
164590 172151 contig of 7562 bp in length
172252 178825 contig of 6574 bp in length

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 79432 contig of 79432 bp in length
* 79433 79532: gap of 100 bp
* 79533 120070: contig of 40538 bp in length
* 120071 120171: gap of 100 bp
* 120171 152785: contig of 32615 bp in length
* 152786 152885: gap of 100 bp
* 152886 164489: contig of 11604 bp in length
* 164490 164589: gap of 100 bp
* 164590 172151: contig of 7562 bp in length
* 172152 172251: gap of 100 bp
* 172252 178825: contig of 6574 bp in length.
```

FEATURES

source

```
1..178825
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-845G3"

1..79432
/notes="assembly_fragment"
79533..120070
/notes="assembly_fragment"
120171..152785
/notes="assembly_fragment"
152886..164489
/notes="assembly_fragment"
164590..172151
/notes="assembly_fragment"
172252..178825
/notes="assembly_fragment clone_end:SP6 vector_side:left"
```

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 178825;
Best Local Similarity 90.5%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 ACTTGTACTGCTTTAAACC 22

Db 24839 ACTTGTACTGCTTTAAACC 24859

RESULT 12

AC101658

LOCUS AC101658 188972 bp DNA linear ROD 30-SEP-2003

DEFINITION Mus musculus chromosome 7, clone RP23-470G18, complete sequence.

AC101658 GI:32490721

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 188972)

Birren, B., Nusbaum, C. and Lander, E.

REFERENCE

JOURNAL

TITLES

AUTHORS

REFERENCE

2 (bases 1 to 188972)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,

Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,

Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeRellano, K., Dewar, K., Diaz, S., Dodge, S., Faro, S.,

Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidys, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heathford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamet, A., Karatas, J., Kells, C., Laroque, K.,

Lamazar, R., Landers, T., Lechoczky, J., Levine, R., Liu, G.,

Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,

Meneis, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strausse, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigg, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 188972)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,

Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choquel, Y.,

Collymore, A., Cook, A., Cooke, P., Corum, B., DeRellano, K.,

Diaz, J.S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Faro, S.,

Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gaidys, S.,

Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,

Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamet, A., Karatas, J., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,

Meldrum, J., Meneis, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,

Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,

O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,

Rachupka, A., Ramasamy, V., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,

Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,

Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,

Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (05-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 188972)
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gaidyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., McDermid, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melidrim, J., Menus, L., Mihova, I., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strubs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (10-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 188972)
Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gaidyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., McDermid, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melidrim, J., Menus, L., Mihova, I., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strubs, M., Talamas, J., Teefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L17165
Center clone name: 470_G_18

Location/Qualifiers
1. 188972
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="7"
/map="7"
/clone="RP23-470G18"
/clone_11b="RPCI-23 Female Mouse BAC"

misc_feature	1. 569 /note="wgs_end_extension clone_end:SP6"
repeat_region	complement(109..148) /rpt_family="MIR"
misc_feature	570..575 /note="clone_boundary clone_end:SP6 site:ECORI"
repeat_region	1189..1209 /rpt_family="AT-rich"
repeat_region	complement(2030..2229) /rpt_family="B3A"
repeat_region	complement(2713..2812) /rpt_family="MER5B"
repeat_region	complement(3596..3709) /rpt_family="MER5A"
repeat_region	3864..3988 /rpt_family="B1F"
repeat_region	3983..4116 /rpt_family="B3"
repeat_region	4175..4275 /rpt_family="(TA)n"
repeat_region	4275..4301 /rpt_family="(GA)n"
repeat_region	4315..4387 /rpt_family="(GGA)n"
repeat_region	4771..4868 /rpt_family="GA-rich"
repeat_region	complement(5908..5967) /rpt_family="MIR"
repeat_region	6172..6459 /rpt_family="B4A"
repeat_region	6471..6516 /rpt_family="(CA)n"
repeat_region	6522..6633 /rpt_family="RSINE1"
repeat_region	9347..9913 /rpt_family="Lx2"
repeat_region	complement(9976..10192) /rpt_family="L2"
repeat_region	10379..10483 /rpt_family="RMR15"
repeat_region	10563..10748 /rpt_family="RMR15"
repeat_region	complement(11784..12143) /rpt_family="MST1"
repeat_region	12299..12854 /rpt_family="RMR19B"
repeat_region	complement(12857..13275) /rpt_family="MTC"
repeat_region	13370..13487 /rpt_family="HERV16"
repeat_region	13515..13667 /rpt_family="B3"
repeat_region	14790..15191 /rpt_family="MTC"
repeat_region	complement(15464..15740) /rpt_family="MTC"
repeat_region	complement(15790..15844) /note="single clone coverage"
repeat_region	15803..16116 /rpt_family="MTE"
repeat_region	complement(16117..16335) /rpt_family="MTR1B"
repeat_region	16336..16427 /rpt_family="MTE"
Query Match	80.9%; Score 17.8; DB 10; Length 188972;
Best Local Similarity	90.5%; Pred. No. 2.1e+02;
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY	2 ATCTGTACTGTCTTAACCC 22

```

Db      179704 ATCTGTACTGCTTTAACCC 179724

RESULT 13
BX470182
LOCUS    BX470182
DEFINITION  BX470182 208244 bp DNA linear HTG 29-JUL-2004
            Danto rerio clone CH211-195P4, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION BX470182.7 GI:50831349
VERSION   HTG; HTGS PHASE1; HTGS ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS  Danto rerio (zebrafish)
SOURCE    Danto rerio
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danto.
REFERENCE 1 (bases 1 to 208244)
AUTHORS   Mashreghi-Mohammadi, M.
TITLE      Direct Submission
JOURNAL    Submitted (28-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgehire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Jul 29, 2004 this sequence version replaced gi:49614046.

COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zc195P4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 207900 bases at least Q40
Consensus quality: 207989 bases at least Q30
Consensus quality: 208016 bases at least Q20
Insert size: 208044; sum-of-contigs
Insert size: 203672; 7.6% error; agarose-fp
Quality coverage: 10.93x in Q20 bases; sum-of-contigs Quality
coverage: 11.26x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 8185: contig of 8185 bp in length
* 8186 8285: gap of 100 bp
* 8286 146916: contig of 138531 bp in length
* 146917 146916: gap of 100 bp
* 146917 208244: contig of 61328 bp in length.
Location/Qualifiers
1. 208244
/organism="Danto rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-195P4"
/clone_1lb="CHORI-211"
1. 8185
/notes="assembly fragment: 00104
fragment chain: 1
clone end: SP6
vector side: left"
8286..146815
/notes="assembly fragment: 03435
fragment chain: 1"
146917..208244
/notes="assembly fragment: 00192
fragment chain: 1"
FEATURES
. source
m1sc_feature
m1sc_feature
m1sc_feature
m1sc_feature

```

```

ORIGIN
Query Match      80.5%; Score 17.8; DB 2; Length 208244;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy      2 ATCTGTACTGCTTTAACCC 22
Db      170912 ATCTGTACTGCTTTAACCC 170932

RESULT 14
AC110680
LOCUS    AC110680
DEFINITION  AC110680.5 213071 bp DNA linear HTG 20-NOV-2002
            Rattus norvegicus clone CH230-250H8, WORKING DRAFT SEQUENCE.
ACCESSION  AC110680.5 GI:25139138
VERSION   HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS  Rattus norvegicus (Norway rat)
SOURCE    Rattus norvegicus
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
1 (bases 1 to 213071)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebreyes, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowls, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwoko, L., O'Connell, G., Olarnpusoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzos, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savary, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,
Shetty, J., Shivaratheya, A., Sisson, I., Sitter, C.D., Smales, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svacek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanzi, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Wortley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhou, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

```

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213071)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
REFERENCE Submitted (15-FEB-2002) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
AUTHORS Baylor Plaza, Houston, TX 77030, USA
TITLE 3 (bases 1 to 213071)
REFERENCE Rat Genome Sequencing Consortium.
AUTHORS Direct Submission
JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 20, 2002 this sequence version replaced gi:213071.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGVF
Center clone name: CH230-250H8
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 202669 bases at least Q40
Consensus quality: 204833 bases at least Q30
Consensus quality: 206266 bases at least Q20
Estimated insert size: 207285; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 213071: contig of 213071 bp in length.
  Location/Qualifiers
    1. .213071
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clone="CH230-250H8"
    1. .2703
      /note="wgs end _extension
      clone end: T7"
      /note="clone_boundary
      clone_end: T7
      site:
      end_sequence: BZ219211"
      complement (206636. .207330)
      /note="clone boundary
      clone_end: SpE
      site:

```

```

misc_feature          end_sequence:18Z219213"
208785..209961
/notes="wgs_end_extension
clone_end:Sp6"
211965..213071
/notes="wgs_end_extension
clone_end:Sp6"

ORIGIN
Query Match      80.9%; Score 17.8; DB 2; Length 213071;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      1 GATCTTGACTGCTCTTAACC 21
|||||
Db      141740 GATCTTGACTGCTCTCAACC 141760

RESULT 15
AC087150/c
LOCUS
DEFINITION
AC087150      223538 bp    DNA          linear   HTG 09-DEC-2000
Mus musculus clone RP23-312B8, WORKING DRAFT SEQUENCE, 20 unordered
pieces.
AC087150
AC087150.1  GI:11610876
HTG; HTGS_PHASE1; HTGS_DRAFT.
Mus musculus (house mouse)
Mus musculus
Eulaliya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 223538)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 223538)
DOE Joint Genome Institute.
Direct Submision
Submitted (09-DEC-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1856712
Center clone name: RPC1-23_312B8
-----
Summary Statistics
Consensus quality: 207168 bases at least Q40
Consensus quality: 211223 bases at least Q30
Consensus quality: 213792 bases at least Q20
Estimated insert size: 210000; agarose-fp estimation
Estimated insert size: 221638; sum-of-ctnigs estimation
Quality coverage: 10.16 in Q20 bases; agarose-fp estimation
Quality coverage: 9.63 in Q20 bases; sum-of-ctnigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1038: contig of 1038 bp in length
1039      1138: gap of unknown length
1139      2206: contig of 1068 bp in length
1207      2306: gap of unknown length
1307      3524: contig of 1218 bp in length
3525      3624: gap of unknown length
3625      5197: contig of 1573 bp in length
5198      5297: gap of unknown length
5298      6391: contig of 1094 bp in length

```



```

* 6392 6491: gap of unknown length
* 6492 8371: contig of 1880 bp in length
* 8372 8372 8471: gap of unknown length
* 8472 8472 9651: contig of 1180 bp in length
* 9652 9652 9751: gap of unknown length
* 9752 11247: contig of 1496 bp in length
* 11348 11348 12711: contig of 1364 bp in length
* 12712 12712 18292: gap of unknown length
* 18292 18292 18392: contig of 5481 bp in length
* 18393 18393 27516: gap of unknown length
* 27517 27517 27516: contig of 9124 bp in length
* 27517 37488: contig of 9872 bp in length
* 37489 37489 52023: gap of unknown length
* 52024 52024 52123: contig of 14435 bp in length
* 52124 52124 69050: gap of unknown length
* 69051 69051 69150: contig of 16927 bp in length
* 69151 91246: contig of 22096 bp in length
* 91247 91247 91346: gap of unknown length
* 91347 106178: contig of 14832 bp in length
* 106179 106278: gap of unknown length
* 106279 131304: contig of 25026 bp in length
* 131305 131404: gap of unknown length
* 131405 153910: contig of 22406 bp in length
* 153911 153911 183927: gap of unknown length
* 153911 183927: contig of 30017 bp in length
* 183928 184028 223538: gap of unknown length
* 184028 223538: contig of 39511 bp in length.

```

FEATURES

```

Source
1..223538
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-312B8"
/location="RP23-312B8"
/clone_lib="RPC1 mouse BAC library 23"

```

ORIGIN

```

Query Match      80.9%; Score 17.8; DB 2; Length 223538;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GATCTGTACTGCTTAACC 21
Db 161710 GATCTTGACTGCTTAGACC 161690

```

Search completed: February 8, 2005, 16:36:51
 Job time : 133.665 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 : Search time 29.1839 Seconds
(without alignments)
3957.220 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22
Sequence: 1 gacctgtactgtcttaaaccc 22

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23sep04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004s:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.2	78.2	363	2	AAX41422 Human sec
C 2	17.2	78.2	454	2	ACH22743 Human adu
C 3	17.2	78.2	457	3	AAC55151 Arabidops
C 4	17.2	78.2	470	6	AB193436 Arabidops
C 5	17.2	78.2	636	3	AAC55036 Arabidops
C 6	17.2	78.2	641	3	AAC53341 Arabidops
C 7	17.2	78.2	101569	12	ADP13274 Renal cel
C 8	17.2	78.2	259202	12	ADQ18492
C 9	16.8	76.4	567	6	ABN78623 Human sof
C 10	16.8	76.4	637	6	ABN61228 Human can
C 11	16.8	76.4	1452	10	ACC60893 Gene seqn
C 12	16.8	76.4	1452	10	ADK62303 Disease t
C 13	16.4	74.5	356	3	AAC29428 Human sec
C 14	16.4	74.5	508	4	ABA60565 Human fce
C 15	16.4	74.5	508	4	AA140454 Probe #91
C 16	16.4	74.5	508	4	AAK34735 Human bon
C 17	16.4	74.5	508	4	AAK08847 Human bra
C 18	16.4	74.5	139257	10	ADC89520 Human COR
C 19	16.2	73.6	199	10	ABK84623 Corn ear-
C 20	16.2	73.6	339	4	AAK53723 Murine tr
C 21	16.2	73.6	473	8	ABZ55713 Aspergill

22	16.2	73.6	499	9	ACL19831	ACL19831 DNA clone
23	16.2	73.6	510	9	ACL19828	ACL19828 DNA clone
24	16.2	73.6	530	9	ACL19827	ACL19827 DNA clone
25	16.2	73.6	534	9	ACL19830	ACL19830 DNA clone
26	16.2	73.6	536	9	ACL19835	ACL19835 DNA clone
27	16.2	73.6	557	3	AAC53403	AAC53403 Arabidops
28	16.2	73.6	565	9	ACL19840	ACL19840 DNA clone
29	16.2	73.6	578	9	ACL19855	ACL19855 DNA clone
30	16.2	73.6	588	2	ADR02268	ADR02268 A. gossyp
31	16.2	73.6	590	9	ACL19847	ACL19847 DNA clone
32	16.2	73.6	591	9	ACL19858	ACL19858 DNA clone
33	16.2	73.6	595	9	ACL19839	ACL19839 DNA clone
34	16.2	73.6	609	9	ACL19854	ACL19854 DNA clone
35	16.2	73.6	618	9	ACL19838	ACL19838 DNA clone
36	16.2	73.6	659	9	ACL19864	ACL19864 DNA clone
37	16.2	73.6	661	9	ACL19860	ACL19860 DNA clone
38	16.2	73.6	662	9	ACL19837	ACL19837 DNA clone
39	16.2	73.6	667	9	ACL19853	ACL19853 DNA clone
40	16.2	73.6	675	9	ACL19836	ACL19836 DNA clone
41	16.2	73.6	682	9	ACL19877	ACL19877 DNA clone
42	16.2	73.6	689	9	ACL19874	ACL19874 DNA clone
43	16.2	73.6	690	9	ACL19867	ACL19867 DNA clone
44	16.2	73.6	691	3	AAC54484	AAC54484 Arabidops
45	16.2	73.6	694	9	ACL19862	ACL19862 DNA clone

ALIGNMENTS

RESULT 1
AAX41422/c
ID AAX41422 standard; cDNA; 363 BP.
XX
AC AAX41422;
XX
DT 22-JUN-1999 (first entry)
XX
DB Human secreted protein 5' EST SEQ ID NO: 81 from WO 9906553.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; antiinflammatory; tumour inhibition; antitumour; de.
XX
OS Homo sapiens.
XX
PN WO9906553-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-1B001237.
XX
PR 01-AUG-1997; 97US-00905051.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
DR WPI: 1999-153783/13.
DR P-PSDB; AAY12564.
XX
PT New nucleic acids encoding human secreted proteins - obtained from cDNA
PT libraries derived from umbilical cord, lymph ganglia, lymphocytes and
XX placental tissue.
XX
PS Claim 1, Page 200, 41pp; English.
XX
CC AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12521 to
CC AAY12668, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid

CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC antipolytic activity, receptor/ligand activity, antiinflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell
XX

SO Sequence 363 BP; 114 A; 79 C; 81 G; 89 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 2; Length 363;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GATCTGTACTGCTTAAACC 22
DB 169 GATCTGTCTGCTTAACTC 148

RESULT 2
ACH22743/c
ID ACH22743 standard; cDNA; 454 BP.
XX
AC ACH22743;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult ovary cDNA #1123.
XX
KW Human; ss; sequencing by hybridisation; SSH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
PI
XX Dramac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 9955; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH22743-ACH20831, whose sequence was
CC determined by the technique of SSH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations

CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX

SO Sequence 454 BP; 149 A; 89 C; 103 G; 113 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 9; Length 454;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GATCTGTACTGCTTAAACC 22
DB 140 GATCTGTCTGCTTAACTC 119

RESULT 3
AAC5151/c
ID AAC5151 standard; DNA; 457 BP.
XX
AC AAC5151;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 80200.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
OS EP1033405-A2.
XX
PN 06-SEP-2000.
XX
PD 25-FEB-2000; 2000EP-00301439.
XX
PF 25-FEB-1999; 99US-0121825P.
XX
PR 05-MAR-1999; 99US-0123180P.
XX
PR 09-MAR-1999; 99US-0123548P.
XX
PR 23-MAR-1999; 99US-0125788P.
XX
PR 25-MAR-1999; 99US-0126264P.
XX
PR 29-MAR-1999; 99US-0126785P.
XX
PR 01-APR-1999; 99US-0127462P.
XX
PR 06-APR-1999; 99US-0128234P.
XX
PR 08-APR-1999; 99US-0128714P.
XX
PR 16-APR-1999; 99US-0128845P.
XX
PR 19-APR-1999; 99US-0130077P.
XX
PR 21-APR-1999; 99US-0130407P.
XX
PR 23-APR-1999; 99US-0130510P.
XX
PR 28-APR-1999; 99US-0130891P.
XX
PR 30-APR-1999; 99US-0132048P.
XX
PR 30-APR-1999; 99US-0132407P.
XX
PR 04-MAY-1999; 99US-0132484P.
XX
PR 05-MAY-1999; 99US-0132485P.
XX
PR 06-MAY-1999; 99US-0132486P.
XX
PR 06-MAY-1999; 99US-0132487P.
XX
PR 07-MAY-1999; 99US-0132863P.
XX
PR 11-MAY-1999; 99US-0134256P.
XX
PR 14-MAY-1999; 99US-0134218P.
XX
PR 14-MAY-1999; 99US-0134221P.
XX
PR 14-MAY-1999; 99US-0134370P.
XX
PR 18-MAY-1999; 99US-0134768P.
XX
PR 19-MAY-1999; 99US-0134941P.
XX
PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140919P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145813P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148665P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151030P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152263P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158323P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.

PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 78.2%; Score 17.2; DB 3; Length 457;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATCTTGACTGCTTTAAACCC 22
DB 189 GATCTTGCTTTGCTATTAACCC 168
RESULT 4
ABL93436
ID ABL93436 standard; cDNA; 470 BP.
AC ABL93436;
XX
XX 10-JUN-2002 (first entry)
XX
DE Arabidopsis thaliana nucleic acid sequence Ref:2027201 SEQ ID NO:201.
XX
XX Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;
KM genetic modification; gene; ss.
XX
OS Arabidopsis thaliana.
XX
XX US2002023280-A1.
XX
XX 21-FEB-2002.
XX
XX 26-JAN-2001; 2001US-00770444.
XX
XX 27-JAN-2000; 2000US-0178502P.
XX
XX (GORL/) GORLACH J.
XX (ANY/) AN Y.
XX (HAM/) HAMILTON C M.
XX (PRIC/) PRICE J L.
XX (RAIN/) RAINES T M.
XX (YUY/) YU Y.
XX (RAME/) RAMEAKA J G.
XX (PAGE/) PAGE A.
XX (MATH/) MATHEW A V.
XX (LEDF/) LEDFORD B L.
XX (WOES/) WOESSNER J P.
XX (HAAS/) HAAS W D.
XX (GARC/) GARCIA C A.
XX (KRIC/) KRICKER M.
XX (SLAT/) SLATER T.
XX (DAVI/) DAVIS K R.
XX (ALLE/) ALLEN K.
XX (HOPE/) HOFFMAN N.
XX (HURB/) HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JT, Raines TM, Yu Y;
XX Rameaka JG, Page A, Mathew AV, Ledford BU, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurban P;
XX WPI; 2002-267486/31.
XX
XX New Arabidopsis thaliana nucleic acid, for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of a protein.
XX
XX Claim 1, SEQ ID NO 201; 44pp; English.
XX
XX The present invention describes an Arabidopsis thaliana nucleic acid (I)
XX comprising a sequence capable of hybridizing under stringent conditions
XX to a sequence (SI) selected from any one of the 999 sequences given in

CC ABL93236 to ABL94234. (1) have insecticide and fungicide activities, and
CC they can be used as protein expression modulators. (1) can be used in
CC identifying homologous or related genes, in producing compositions that
CC modulate the expression or function of their encoded proteins, mapping
CC functional regions of the protein, and in studying associated
CC physiological pathways. (1) can also be used: (1) for the genetic
CC manipulation of cells, particularly plant cells; (2) in screening assays
CC of various plant strains to determine the strains that are best capable
CC of withstanding a particular disease or environmental stress; (3) for
CC enhancing or inhibiting production of a biosynthetic product in a plant;
CC (4) as probes in mapping and in diagnosis, in genetic modification and
CC for screening purposes, to generate additional copies of the nucleic
CC acids, to generate ribozymes or antisense oligonucleotides, and as single
CC stranded DNA probes or as triple-strand forming oligonucleotides; and (5)
CC for generating genetically modified transgenic organisms. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site
XX
SQ Sequence 470 BP; 116 A; 107 C; 95 G; 152 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 6; Length 470;
Best Local Similarity 86.4%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATCTTGACTGCTTTAAACCC 22
DB 292 GATCTTGCTTTGCTATTAACCC 313
RESULT 5
AAC5036/C
ID AAC5036 standard; DNA; 636 BP.
XX
XX AAC5036;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 79959.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135174P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140921P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 27-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151330P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 16-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 78.2%; Score 17.2; DB 3; Length 636;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATCTGTACTGCTTAACCC 22
Db 363 GATCTGTCTTGTGATTAACCC 342

RESULT 6
AAC53341/c
ID AAC53341 standard; DNA; 641 BP.
XX AAC53341;
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 74193.
DE
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway; metabolic pathway;
KW Promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134321P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145090P.
PR 22-JUL-1999; 99US-0145091P.
PR 23-JUL-1999; 99US-0145122P.
PR 23-JUL-1999; 99US-0145148P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.


```

PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 24-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156586P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157665P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159332P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

```

```

PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      78.2%; Score 17.2; DB 3; Length 641;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Cy 1 GATCTGTACTGCTTAACCC 22
Db 365 GATCTGTTTGTCAATAACCC 344

RESULT 7
ADP13274/c
ID ADP13274 standard; DNA; 101569 BP.
XX
AC ADP13274;
XX
DT 26-AUG-2004 (first entry)
XX
DE Renal cell carcinoma differentially expressed gene #10.
XX
KW de; diagnosis; non-blood disease; solid tumor; gene expression;
KW peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW head/neck cancer; differential expression.
XX
OS Homo sapiens.
XX
EN WO2004048933-A2.
XX
PD 10-JUN-2004.
XX
PF 21-NOV-2003; 2003MO-US037481.
XX
PR 21-NOV-2002; 2002US-0427982P.
PR 03-APR-2003; 2003US-0459782P.
XX
PA (AMHP ) WYETH.
PA (TWIN/) TWINE N C.
PA (BURC/) BURCZYNSKI M E.
PA (TREP/) TREPICCHIO W L.
PA (DORN/) DORNER A.
PA (STOV/) STOVER J A.
PA (SLOW/) SLOWI D K.
XX
PI Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
PI Slowi DK;
XX
DR WPI; 2004-460799/43.
XX
PT diagnosing non-blood disease such as solid tumor, involves comparing
PT differential expression profile of specific genes in peripheral blood
PT sample of subject with reference expression profile of specific genes.
XX
XX Disclosure; SEQ ID NO 10; 350bp; English.
XX
CC The invention relate to a method of diagnosing (M1) non-blood disease
CC such as solid tumor by providing peripheral blood sample of human having
CC non-blood disease, and comparing an expression profile of specific genes
CC in the peripheral blood sample to reference expression profile of the
CC genes, where each of the genes is differentially expressed in peripheral
CC blood mononuclear cells (PBMCs) of patients having the disease as
CC compared to PBMCs of normal humans. The method is useful for diagnosing
CC non-blood disease such as solid tumor. The solid tumor is chosen from
CC renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC sample is a whole blood sample (claimed). (M1) is useful for identifying
CC genes that are differentially expressed in peripheral blood samples
CC isolated at different stages of progression, development or treatment of

```

CC RCC and/or other solid tumors. This sequence corresponds to a gene that
CC is differentially expressed and detected by the method of the invention.
CC (Note: this sequence is not given as part of the printed specification
CC but was obtained from WIFO in electronic format at
CC ftp.wifo./pub/published_pct_sequences).

XX
SQ Sequence 101569 BP; 27935 A; 22428 C; 22623 G; 28582 T; 0 U; 1 Other;

Query Match 78.2%; Score 17.2; DB 12; Length 101569;
Best Local Similarity 86.4%; Pred. No. 3e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATCTTGTACTGTCTTAAACCC 22
Db 55369 GATCTTGTACTGTCTTAAACCC 55348

RESULT 8
ADQ18492/c
ID ADQ18492 standard; DNA; 259202 BP.

XX
AC ADQ18492;

XX
DT 26-AUG-2004 (first entry)

XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.

XX
KM soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM da.

XX
OS Homo sapiens.

XX
PN MO2004048938-A2.

XX
PD 10-JUN-2004.

XX
PF 26-NOV-2003; 2003WO-US038193.

XX
PR 26-NOV-2002; 2002US-0429739P.

XX
PA (PROT-) PROTEIN DESIGN LABS INC.

XX
PI Aziz N, Ginsburg WM, Zlotnick A;

XX
DR WPI; 2004-441208/41.

XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX
PS Example 2; SEQ ID NO 1311; 210pp; English.

XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX
SQ Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 12; Length 259202;
Best Local Similarity 86.4%; Pred. No. 3.2e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATCTTGTACTGTCTTAAACCC 22

Db 70030 GATCTTGTACTGTCTTAAACCC 70009

RESULT 9

ABN78623/c
ID ABN78623 standard; cDNA; 567 BP.

XX
AC ABN78623;

XX
DT 08-JUL-2002 (first entry)

XX
DE Human synthase-like ORF3570 cDNA, SEQ ID NO:7139.

XX
KM Human; ORF, open reading frame; ORFX; drug screening; diagnosis;
KM disease monitoring; cytokine; cell proliferation; cell differentiation;
KM immune modulation; haematopoiesis regulation; tissue growth;

XX
KM angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KM thrombolytic; tumour inhibition; bodily characteristics; fertility;

XX
KM behaviour; cancer; proliferative disorder; neurological disorder;
KM cardiovascular disease; immune system disorder; organ transplantation;

XX
KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KM hypothyroidism; cholesterol ester storage disease; infection; vulnery;

XX
KM vasotropic; antiporatic; antidiabetic; cyrostatic; nootropic;
KM neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;

XX
KM cardiant; hypotensive; antihyroid; antiinflammatory; immunomodulator;
KM dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.

XX
OS Homo sapiens.

XX
PN MO200190366-A2.

XX
PD 29-NOV-2001.

XX
PF 24-MAY-2001; 2001WO-US017076.

XX
PR 24-MAY-2000; 2000US-0206690P.

XX
PA (CUPRA-) CUPAGEN CORP.

XX
PI Leach MD, Shinkels RA;

XX
DR WPI; 2002-106200/14.

XX
PT P-PSDB; ABP34597.

XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.

XX
PS Claim 1; Page 2031; 2508pp; English.

XX
CC Sequences ABP1028-ABP3561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antifective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,

CC neurological disorders such as epilepsy and Alzheimer's disease.
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases

XX
SQ Sequence 567 BP, 146 A; 97 C; 133 G; 191 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 6; Length 567;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGCTTAAACC 22
DB 303 TCTGTATGTCTGTAACCC 284

RESULT 10
ABN61229/c
ID ABN61229 standard; cDNA; 637 BP.
XX
AC ABN61229;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 1196.
XX
KM Human; cytostatic; gene expression; gene mapping; tissue profiling;
XX gene therapy; cancer; tumour; gene; 88.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US025840.
XX
PR 16-AUG-2000; 2000US-0226326P.
XX
PA (CHIR) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
PI Sacobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
XX Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX WPI; 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
XX expressed genes correlated with a cancerous state of a mammalian cell,
XX and inhibiting tumor growth.

PS Claim 1; SEQ ID NO 1196; 883pp + Sequence Listing; English.

CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 637 BP, 265 A; 84 C; 122 G; 166 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 6; Length 637;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGCTTAAACC 22
DB 543 TCTGTACTGCTTCAACC 524

RESULT 11
ACC60893/c
ID ACC60893 standard; DNA; 1452 BP.
XX
AC ACC60893;
XX
DT 20-JUN-2003 (first entry)
XX
DE Gene sequence #SEQ ID 568.
XX
KM Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PF 20-DEC-2001; 2001EP-00130253.
XX
PR 15-MAY-2001; 2001EP-0011774.
XX
PA (CELL-) CELLZONE AG.
XX
PI Bauer A, Grandi P, Krause R, Kruse UD, Kuester BD;
XX Marzloch M, Schultz JD, Supercil-Furga GD;
XX P-PSDB; ABR52851.
XX
DR WPI; 2003-250078/25.
XX
PT New isolated protein complexes useful for diagnosing a disease or
XX disorder, or as a target for an active agent of a pharmaceutical,
XX preferably a drug target in the treatment or prevention of disease or
XX disorder.

PS Disclosure; SEQ ID NO 568; 17pp + Sequence Listing; English.

CC The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM

XX
SQ Sequence 1452 BP, 489 A; 226 C; 351 G; 386 T; 0 U; 0 Other;

Query Match 76.4%; Score 16.8; DB 10; Length 1452;
Best Local Similarity 90.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGCTTAAACC 22
DB 741 TCTGTATGTCTGTAACCC 722

RESULT 12
 ADK62303/C
 ID ADK62303 standard; DNA; 1452 BP.
 XX
 AC ADK62303;
 XX
 DT 06-MAY-2004 (first entry)
 DE Disease treating protein complex-derived gene #277.
 XX
 KW protein complex; drug target; diagnosis; gene; ds.
 XX
 OS unidentified.
 XX
 PN EPI338608-A2.
 XX
 PD 27-AUG-2003.
 XX
 PF 20-DEC-2002; 2002EP-00102902.
 XX
 PR 20-DEC-2001; 2001EP-00130253.
 XX
 PA (CELL-) CELLZONE AG.
 XX
 PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;
 PI Marzloch M, Grandt P, Krause R, Kruse U, Merino A, Bauch A;
 PI Michon A, Leutwein C, Rick J;
 DR WPI; 2003-638460/61.
 DR P-PSDB; ADK62302.
 XX
 PT New proteins and protein complexes from eukaryotes, useful as targets in
 PT drug screening, or in diagnosing or screening for the presence of a
 PT disease or disorder, or a predisposition for developing a disease or
 PT disorder in a subject.
 XX
 PS Disclosure; SEQ ID NO 554; 13bp; English.
 XX
 CC The invention relates to novel protein complexes comprising a first and a
 CC second protein, or its derivative, fragment, homologue or variant. The
 CC proteins are selected from given protein complexes, which are not defined
 CC in the specification. The variants are encoded by nucleic acids that
 CC hybridize to the nucleic acids encoding the proteins under low stringency
 CC conditions. The protein complexes are useful as targets for an active
 CC agent of a pharmaceutical. These protein complexes are particularly
 CC useful as drugs targets for the treatment or preventing of a disease or
 CC disorder. The complexes and methods above are useful in diagnosing or
 CC screening for the presence of a disease or disorder or a predisposition
 CC for developing a disease or disorder in a subject. These are also useful
 CC in screening for a drug for treatment or prevention of a disease or
 CC disorder. The molecule that modulates the amount, activity or protein
 CC components of the complex is useful for the manufacture of a medicament
 CC for the treatment or prevention of a disease or disorder. This sequence
 CC corresponds to a gene of the invention. (Note: the sequence data for this
 CC patent did not form part of the printed specification but was obtained
 CC from the EPO in electronic format).
 XX
 SQ Sequence 1452 BP; 489 A; 226 C; 351 G; 386 T; 0 U; 0 Other;
 XX
 Query Match 76.4%; Score 16.8; DB 10; Length 1452;
 Best Local Similarity 90.0%; Pred. No. 3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TCTGTACTGCTTAAACC 22
 |||||
 DB 741 TCTGTATGTCGTAACCC 722

RESULT 13
 AAC29428
 ID AAC29428 standard; cDNA; 356 BP.
 XX
 AC AAC29428;
 PD

XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 33503.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-00200610.
 XX
 PR 26-FEB-1999; 99US-0122487P.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI; 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 PS Claim 1, SEQ ID NO 33503; 71bp + Sequence listing; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX
 SQ Sequence 356 BP; 93 A; 81 C; 52 G; 130 T; 0 U; 0 Other;
 XX
 Query Match 74.5%; Score 16.4; DB 3; Length 356;
 Best Local Similarity 94.4%; Pred. No. 4.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 TCTGTACTGCTTAAAC 20
 |||||
 DB 332 TCTTTACTGCTTAAAC 349

RESULT 14
 ABA60565
 ID ABA60565 standard; DNA; 508 BP.
 XX
 AC ABA60565;
 XX
 DT 01-FEB-2002 (first entry)
 DE Human foetal liver single exon nucleic acid probe #8870.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.

```

XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 8870; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 508 BP; 131 A; 118 C; 95 G; 164 T; 0 U; 0 Other;

Query Match          74.5%; Score 16.4; DB 4; Length 508;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TCTTGACTGCTTTAAAC 20
DB 255 TCTTTACTGCTTTAAAC 272

RESULT 15
AA140454
ID AA140454 standard; DNA; 508 BP.
XX AC AA140454;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #9140 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
```

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 25; SEQ ID NO 9140; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX SQ Sequence 508 BP; 131 A; 118 C; 95 G; 164 T; 0 U; 0 Other;

Query Match          74.5%; Score 16.4; DB 4; Length 508;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TCTTGACTGCTTTAAAC 20
DB 255 TCTTTACTGCTTTAAAC 272

Search completed: February 8, 2005, 13:56:56
Job time : 34.1839 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 5.94179 Seconds
(Without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22
Sequence: 1 gatctgtactgtcttaacc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued, Patents NA: *
1: /cgn2_6/prodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/prodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	78.2	479	4 US-09-621-976-3622	Sequence 3622, App
2	16.4	74.5	356	4 US-09-513-999C-33503	Sequence 33503, A
3	16.2	73.6	199	4 US-09-313-394A-3083	Sequence 3083, Ap
4	16.2	73.6	279	4 US-09-248-796A-8060	Sequence 8060, Ap
5	16.2	73.6	588	3 US-08-998-416-960	Sequence 960, App
6	16.2	73.6	5554	4 US-09-815-923-1	Sequence 1, Appl
7	16	72.7	741	4 US-09-543-681A-4040	Sequence 4040, App
8	15.8	71.8	567	4 US-09-107-532A-618	Sequence 618, App
9	15.6	70.9	346	4 US-09-513-999C-10254	Sequence 10254, A
10	15.6	70.9	1039	4 US-09-566-921-2	Sequence 2, Appl
11	15.6	70.9	1980	4 US-09-543-681A-576	Sequence 576, App
12	15.4	70.0	1251	4 US-10-164-595-25	Sequence 25, Appl
13	15.4	70.0	1341	4 US-08-956-171E-448	Sequence 448, App
14	15.4	70.0	1341	4 US-08-781-986A-448	Sequence 448, App
15	15.4	70.0	14770	4 US-09-230-132-30	Sequence 448, App
16	15.4	70.0	392000	4 US-10-027-983-11	Sequence 11, Appl
17	15.2	69.1	306	2 US-08-634-797-16	Sequence 16, Appl
18	15.2	69.1	306	2 US-08-634-797-16	Sequence 16, Appl
19	15.2	69.1	391	4 US-09-513-999C-26343	Sequence 26343, A
20	15.2	69.1	524	4 US-09-270-767-11178	Sequence 11178, A
21	15.2	69.1	528	4 US-09-583-110-1687	Sequence 1687, Ap
22	15.2	69.1	660	4 US-09-248-796A-4830	Sequence 4830, Ap
23	15.2	69.1	1020	4 US-09-107-532A-1860	Sequence 1860, Ap
24	15.2	69.1	1146	4 US-09-489-039A-985	Sequence 985, App
25	15.2	69.1	1296	4 US-08-714-741-43	Sequence 43, Appl
26	15.2	69.1	1587	4 US-09-248-796A-3998	Sequence 3998, Ap
27	15.2	69.1	2059	4 US-08-714-741-45	Sequence 45, Appl

ALIGNMENTS

C 28	15.2	69.1	3023	3 US-09-308-022-4	Sequence 4, Appl
C 29	15.2	69.1	3222	4 US-08-714-741-39	Sequence 39, Appl
C 30	15.2	69.1	1664976	4 US-08-916-421B-1	Sequence 1, Appl
C 31	15.2	69.1	1664976	4 US-09-692-570-1	Sequence 1, Appl
C 32	15	68.2	10419	4 US-09-408-020-3	Sequence 3, Appl
C 33	15	68.2	42432	4 US-09-408-020-2	Sequence 2, Appl
C 34	14.8	67.3	263	4 US-09-513-999C-20021	Sequence 20021, A
C 35	14.8	67.3	375	4 US-09-248-796A-12298	Sequence 12298, A
C 36	14.8	67.3	423	1 US-08-470-179-150	Sequence 150, App
C 37	14.8	67.3	1189	4 US-09-799-451-709	Sequence 709, App
C 38	14.8	67.3	1833	4 US-09-248-796A-2927	Sequence 2927, App
C 39	14.8	67.3	2084	4 US-09-222-938A-9	Sequence 9, Appl
C 40	14.8	67.3	8054	4 US-09-581-909-2	Sequence 2, Appl
C 41	14.8	67.3	28626	4 US-09-586-002-15	Sequence 15, Appl
C 42	14.8	67.3	1830121	4 US-09-557-884-1	Sequence 1, Appl
C 43	14.8	67.3	1830121	4 US-09-643-990A-1	Sequence 1, Appl
C 44	14.8	67.3	1830121	4 US-10-329-960-1	Sequence 1, Appl
C 45	14.6	66.4	347	4 US-09-621-976-12653	Sequence 12653, A

RESULT 1

US-09-621-976-3622
Sequence 3622, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 3622
LENGTH: 479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 133..282
US-09-621-976-3622

Query Match

Best local Similarity 86.4%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGTCTTAACCC 22
DB 421 GATCTGTACTGTCTTAACCC 442

RESULT 2

US-09-513-999C-33503
Sequence 33503, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59 US2, REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 33503

LENGTH: 356
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-33503

Query Match 74.5%; Score 16.4; DB 4; Length 356;
Best Local Similarity 94.4%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TCTTGTACTGCTTAAC 20
DB 332 TCTTGTACTGCTTAAC 349

RESULT 3
US-09-313-294A-3083
Sequence 3083, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Laligudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 3083
LENGTH: 199
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6476212 700610840H1
NAME/KEY: unsure
LOCATION: 6, 86
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-3083

Query Match 73.6%; Score 16.2; DB 4; Length 199;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAACCC 22
DB 8 ATCTGTACTGCTTAACCC 28

RESULT 4
US-09-248-796A-8060
Sequence 8060, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 8060
LENGTH: 279
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-8060

Query Match 73.6%; Score 16.2; DB 4; Length 279;
Best Local Similarity 85.7%; Pred. No. 53;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAACCC 22
DB 43 ATCTGTCTCTGTATTAACCC 63

RESULT 5
US-08-998-416-960
Sequence 960, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264ch Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 960:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PLAG1592UP
US-08-998-416-960

Query Match 73.6%; Score 16.2; DB 3; Length 588;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAACCC 22
DB 157 ATCATGTACTGCTTAACGC 177

RESULT 6
US-09-815-923-1
Sequence 1, Application US/09815923
Patent No. 6787642
GENERAL INFORMATION:


```

; APPLICANT: Gill, Sarjeet S.
; APPLICANT: Rose, Linda S.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. 6787642e1
; FILE REFERENCE: 023070-093800US
; CURRENT APPLICATION NUMBER: US/09/815,923
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5554
; TYPE: DNA
; ORGANISM: Manduca sexta
; OTHER INFORMATION: vesicular acetylcholine transporter
US-09-815-923-1

Query Match      73.6%; Score 16.2; DB 4; Length 5554;
Best Local Similarity 85.7%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTGTACTGTCTTAACC 21
Db      2219 GATCTGTACTGTCTTAACC 2239

RESULT 7
US-09-543-681A-4040
; Sequence 4040, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4040
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-4040

Query Match      72.7%; Score 16; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 GTACTGTCTTAACC 22
Db      356 GTACTGTCTTAACC 371

RESULT 8
US-09-107-532A-618/C
; Sequence 618, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 618:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (b) LOCATION 1...567
; SEQUENCE DESCRIPTION: SEQ ID NO: 618:
US-09-107-532A-618

Query Match      71.8%; Score 15.8; DB 4; Length 567;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 TCTTGTACTGTCTTAACC 21
Db      474 TCTTGTACTGTCTTAACC 456

RESULT 9
US-09-513-999C-10254/C
; Sequence 10254, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59,US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 10254
; LENGTH: 346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 125
; OTHER INFORMATION: n=a, g, c or t
US-09-513-999C-10254

Query Match      70.9%; Score 15.6; DB 4; Length 346;
```

Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22

Db 207 GATCTGAACAGCTCCACCC 186

RESULT 10

US-09-566-921-2/c
; Sequence 2, Application US/09566921
; Patent No. 668288
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1039
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 668288 998433.2
US-09-566-921-2

Query Match 70.9%; Score 15.6; DB 4; Length 1039;
Best Local Similarity 81.8%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22

Db 648 GATCTGAACAGCTCCACCC 627

RESULT 11

US-09-543-681A-576/c
; Sequence 576, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 576
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-576

Query Match 70.9%; Score 15.6; DB 4; Length 1980;
Best Local Similarity 81.8%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACCC 22

Db 280 GATCTGTATGCTTAATCC 259

RESULT 12

US-10-164-595-25
; Sequence 25, Application US/10164595
; Patent No. 6657054

; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (60)..(1136)
; OTHER INFORMATION:
US-10-164-595-25

Query Match 70.0%; Score 15.4; DB 4; Length 1251;
Best Local Similarity 94.1%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTA 18

Db 1167 ATCTGTACTGCTTA 1183

RESULT 13

US-08-956-171E-448
; Sequence 448, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; Gal H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 448:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

```

;
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 448:
US-08-956-171E-448

Query Match
Best Local Similarity 94.0%; Score 15.4; DB 4; Length 1341;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTTGACTGCTTTAA 18
|||||
Db 1220 ATCTTGACTGCTTTAA 1236

RESULT 14
US-08-781-986A-448
; Sequence 448, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 525
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 448:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-448

Query Match
Best Local Similarity 94.0%; Score 15.4; DB 4; Length 1341;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTTGACTGCTTTAA 18
|||||
Db 1220 ATCTTGACTGCTTTAA 1236

RESULT 15
US-09-220-132-30/C
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; TITLE OF INVENTION: OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 0733-074001

```

```

? CURRENT APPLICATION NUMBER: US/09/220,132
? CURRENT FILING DATE: 1998-12-23
? PRIOR APPLICATION NUMBER: US 60/079,303
? PRIOR FILING DATE: 1998-03-25
? PRIOR APPLICATION NUMBER: US 60/068,821
? PRIOR FILING DATE: 1997-12-24
? NUMBER OF SEQ ID NOS: 191
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 30
? LENGTH: 14770
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-220-132-30

Query Match          70.0%; Score 15.4; DB 4; Length 14770;
Best Local Similarity 94.1%; Pred.No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 TCTGTACTGCTTAA 19
        ||| ||||| ||||| |||
Db       9572 TCTGTACTGCTTACA 9556

```

Search completed: February 8, 2005, 22:25:57
Job time : 13.0529 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 29.6449 Seconds

(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22
Sequence: 1 gactctgtactgtcttaacc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	78.2	220	US-10-674-124A-21842	Sequence 21842, A
2	17.2	78.2	454	US-09-918-995-9955	Sequence 9955, Ap
3	17.2	78.2	470	US-09-770-444-201	Sequence 201, App
4	17.2	78.2	600	US-10-027-632-287049	Sequence 287049, A
5	17.2	78.2	600	US-10-027-632-287050	Sequence 287050, A
6	17.2	78.2	600	US-10-027-632-287049	Sequence 287049, A
7	17.2	78.2	600	US-10-027-632-287050	Sequence 287050, A
8	17.2	78.2	3355	US-10-027-632-114078	Sequence 114078, A
9	17.2	78.2	3355	US-10-027-632-114078	Sequence 114078, A
10	17.2	78.2	39079	US-10-322-281-832	Sequence 832, App
11	17.2	78.2	101569	US-10-717-597-10	Sequence 10, Appl
12	17.2	78.2	259202	US-10-723-860-1311	Sequence 1311, Ap

C 13	17	77.3	77834	13	US-10-087-192-343	Sequence 343, App
C 14	16.8	76.4	567	11	US-09-864-408A-7139	Sequence 7139, Ap
C 15	16.8	76.4	568	18	US-10-425-115-149520	Sequence 149520, A
C 16	16.8	76.4	835	17	US-10-437-963-93764	Sequence 93764, A
C 17	16.8	76.4	1301	17	US-10-767-701-9447	Sequence 9447, Ap
C 18	16.8	76.4	2568	17	US-10-437-963-93762	Sequence 93762, A
C 19	16.4	74.5	508	9	US-09-864-761-12341	Sequence 12341, A
C 20	16.4	74.5	593	13	US-10-027-632-47128	Sequence 47128, A
C 21	16.4	74.5	593	15	US-10-027-632-47128	Sequence 47128, A
C 22	16.4	74.5	625	16	US-10-424-599-117287	Sequence 117287, A
C 23	16.4	74.5	777	13	US-10-027-632-21575	Sequence 21575, A
C 24	16.4	74.5	777	15	US-10-027-632-21575	Sequence 21575, A
C 25	16.4	74.5	139257	10	US-09-920-671-11	Sequence 11, Appl
C 26	16.2	73.6	281	16	US-10-424-599-68086	Sequence 68086, A
C 27	16.2	73.6	339	17	US-10-469-285-288	Sequence 288, App
C 28	16.2	73.6	1039	17	US-10-767-701-15302	Sequence 15302, A
C 29	16.2	73.6	1661	18	US-10-739-930-5539	Sequence 5539, Ap
C 30	16.2	73.6	1760	17	US-10-767-701-15450	Sequence 15450, A
C 31	16.2	73.6	1777	16	US-10-425-114-31265	Sequence 31265, A
C 32	16.2	73.6	1787	16	US-10-425-114-24165	Sequence 24165, A
C 33	16.2	73.6	2000	9	US-09-938-842A-4001	Sequence 4001, Ap
C 34	16.2	73.6	2000	11	US-09-938-842A-4001	Sequence 4001, Ap
C 35	16.2	73.6	2515	18	US-10-425-115-58255	Sequence 58255, A
C 36	16.2	73.6	3239	13	US-10-027-632-114978	Sequence 114978, A
C 37	16.2	73.6	3239	15	US-10-027-632-114978	Sequence 114978, A
C 38	16.2	73.6	5554	9	US-09-815-923-1	Sequence 1, Appl
C 39	16.2	73.6	14759	10	US-09-764-891-8344	Sequence 8344, Ap
C 40	16.2	73.6	14759	15	US-10-205-428-777	Sequence 777, App
C 41	16.2	73.6	17146	9	US-09-764-877-3850	Sequence 3850, Ap
C 42	16.2	73.6	17146	16	US-10-242-515-3850	Sequence 3850, Ap
C 43	16.2	73.6	30568	9	US-09-764-877-3851	Sequence 3851, Ap
C 44	16.2	73.6	30568	16	US-10-242-515-3851	Sequence 3851, Ap
C 45	16.2	73.6	49031	17	US-10-322-281-523	Sequence 523, App

ALIGNMENTS

RESULT 1
US-10-674-124A-21842
; Sequence 21842, Application US/10674124A
; Publication No. US2004019797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hidetoshi
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; FILE REFERENCE: ORIN-003CIP
; CURRENT APPLICATION NUMBER: US/10/674,124A
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 21842
; LENGTH: 220
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Z66933
; OTHER INFORMATION: Located on chromosome 16
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; OTHER INFORMATION: sequence : 9358990

```
FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 5631
US-10-674-124A-21842

Query Match      78.2%; Score 17.2; DB 18; Length 220;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTTGACTGCTTAACCC 22
Db      148 GATCTTGAAATGTTGAACCC 169

RESULT 2
US-09-918-995-9955/c
; Sequence 9955, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-736
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ. ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9955
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-9955

Query Match      78.2%; Score 17.2; DB 10; Length 454;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTTGACTGCTTAACCC 22
Db      140 GATCTTGTTGCTTTAACTC 119

RESULT 3
US-09-770-444-201
; Sequence 201, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesener, Jeffrey P.
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Huben, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
```

```
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ. ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-201

Query Match      78.2%; Score 17.2; DB 9; Length 470;
Best Local Similarity 86.4%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTTGACTGCTTAACCC 22
Db      292 GATCTTGTTTGTCAATAACCC 313

RESULT 4
US-10-027-632-287049
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ. ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

Query Match      78.2%; Score 17.2; DB 13; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 GATCTTGACTGCTTAACCC 22
Db      99 GATCTCACTGTTTAACCC 120

RESULT 5
US-10-027-632-287050
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
```

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match      78.2% Score 17.2; DB 13; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCACTACTGTTTAAACC 120
```

```

RESULT 6
US-10-027-632-287049
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      78.2% Score 17.2; DB 15; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCACTACTGTTTAAACC 120
```

```

RESULT 7
US-10-027-632-287050
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
```

```
Query Match      78.2% Score 17.2; DB 15; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 GATCTGTACTGCTTTAAACC 22
      ||||| ||||| ||||| |||||
DB      99 GATCTCACTACTGTTTAAACC 120
```

```

RESULT 8
US-10-027-632-114078
; Sequence 114078, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114078
; LENGTH: 3355
; TYPE: DNA
; ORGANISM: Human
```

US-10-027-632-114078

Query Match 78.2%; Score 17.2; DB 13; Length 3355;
Best Local Similarity 86.4%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTAACCC 22
Db 2696 GGTCTGTACTGCTGAATCCC 2717

RESULT 9

US-10-027-632-114078
; Sequence 114078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/2218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114078
; LENGTH: 3355
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114078

Query Match 78.2%; Score 17.2; DB 15; Length 3355;
Best Local Similarity 86.4%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTAACCC 22
Db 2696 GGTCTGTACTGCTGAATCCC 2717

RESULT 10

US-10-322-281-832/c
; Sequence 832, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 832
; LENGTH: 39079
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-322-281-832

Query Match 78.2%; Score 17.2; DB 17; Length 39079;

Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTAACCC 22
Db 37200 GATCTGAACGTGCTTAATCCC 37179

RESULT 11

US-10-717-597-10/c
; Sequence 10, Application US/10717597
; Publication No. US20040110221A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Burczynski, Michael E.
; APPLICANT: Twine, Natalie C.
; APPLICANT: Dornier, Andrew J.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Stonim, Donna K.
; APPLICANT: Stover, Jennifer A.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
; FILE REFERENCE: AM101080L
; CURRENT APPLICATION NUMBER: US/10/717,597
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US 60/459,782
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/427,982
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 101569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (22955)..(22955)
; OTHER INFORMATION: No residue exists at "n" position (position 22955) in the
; OTHER INFORMATION: corresponding human genome sequence (ET82 gene) in the Entrez
US-10-717-597-10

Query Match 78.2%; Score 17.2; DB 17; Length 101569;
Best Local Similarity 86.4%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTAACCC 22
Db 55369 GATCTGAACGTGCTTAATCCC 55348

RESULT 12

US-10-723-860-1311/c
; Sequence 1311, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natascha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1311
; LENGTH: 259202
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 78.2%; Score 17.2; DB 18; Length 259202;
Best Local Similarity 86.4%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGTACTGTCTTAAACC 22
DB 70030 GATCTCATCTGTCTTAAACC 70009

RESULT 13
US-10-087-192-343/C
; Sequence 343, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 343
; LENGTH: 77834
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(77834)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-343

Query Match 77.3%; Score 17; DB 13; Length 77834;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAA 19
DB 64684 TCTTGACTGTCTTAA 64668

RESULT 14
US-09-864-408A-7139/C
; Sequence 7139, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7139
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-7139

Query Match 76.4%; Score 16.8; DB 11; Length 567;
Best Local Similarity 90.0%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAAACC 22

DB 303 TCTTGACTGTCTTAAACC 284

RESULT 15
US-10-425-115-149520
; Sequence 149520, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 149520
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(568)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_6788C.1
US-10-425-115-149520

Query Match 76.4%; Score 16.8; DB 18; Length 568;
Best Local Similarity 90.0%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TCTTGACTGTCTTAAACC 22
DB 36 TCTTGACTGTCTTAAACC 55

Search completed: February 9, 2005, 06:50:56
Job time : 33.6449 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 262.041 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-5

Perfect score: 22
Sequence: 1 gactctgactgctttaaacc 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: gb_ests1:*
2: gb_ests2:*
3: gb_hnc:*
4: gb_ests3:*
5: gb_ests4:*
6: gb_ests5:*
7: gb_ests6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	85.5	300	1	AJ461766
C 2	18.8	85.5	423	1	AV692067
C 3	17.8	80.9	505	1	AU003892
C 4	17.8	80.9	606	8	B2419788
C 5	17.8	80.9	711	9	AG402501
C 6	17.8	80.9	799	4	CC638234
C 7	17.8	80.9	804	4	CG289973
C 8	17.8	80.9	885	9	CG320249
C 9	17.8	80.9	932	9	CG324260
C 10	17.4	79.1	500	5	BP185852
C 11	17.4	79.1	685	9	CE153858
C 12	17.4	79.1	700	9	CG922134
C 13	17.4	79.1	740	9	AG452989
C 14	17.2	78.2	242	6	CB261729
C 15	17.2	78.2	300	1	AU099200
C 16	17.2	78.2	310	1	AA380417
C 17	17.2	78.2	305	1	AA227042
C 18	17.2	78.2	321	6	CD529079
C 19	17.2	78.2	346	8	AZ510699
C 20	17.2	78.2	360	6	CS1188
C 21	17.2	78.2	360	7	D67108
C 22	17.2	78.2	369	5	BP654637
C 23	17.2	78.2	376	1	A1870075
C 24	17.2	78.2	376	5	BP649093

25	17.2	78.2	384	8	A0170514
26	17.2	78.2	388	1	AV799434
27	17.2	78.2	398	4	BM755387
28	17.2	78.2	402	1	AV789942
29	17.2	78.2	402	7	CK120300
30	17.2	78.2	403	5	BP647245
31	17.2	78.2	406	1	AV815144
32	17.2	78.2	411	1	AA597625
33	17.2	78.2	418	1	AV788538
34	17.2	78.2	421	6	CB263536
35	17.2	78.2	423	7	CK118569
36	17.2	78.2	427	7	R90061
37	17.2	78.2	431	5	BP626409
38	17.2	78.2	432	3	CN80A6FL
39	17.2	78.2	434	3	CN80A6FL
40	17.2	78.2	437	6	CB265220
41	17.2	78.2	438	7	CK121985
42	17.2	78.2	440	8	AQ050661
43	17.2	78.2	441	7	CK118716
44	17.2	78.2	444	7	CK117689
45	17.2	78.2	444	7	CK119024

ALIGNMENTS

RESULT 1
AJ461766/c 300 bp mRNA linear EST 24-MAY-2002
DEFINITION AJ461766 S00002 Hordeum vulgare subsp. vulgare cDNA clone
VERSION S0000200021A12F1, mRNA sequence.
ACCESSION AJ461766
KEYWORDS GI:21060686
SOURCE
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 300)

REFERENCE
Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.
Barley EST's
TITLE
JOURNAL
COMMENT
Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O. Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES
source 1..300
Location/Qualifiers
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Saana"
/sub_species="vulgare"
/db_xref="taxon:112509"
/dev_stage="Embryo"
/clone_lib="S00002"
/note="1 day after pollination"

ORIGIN

Query Match 85.5%; Score 18.8; DB 1; Length 300;
Best Local Similarity 90.9%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAAACC 22
|||||
DB 212 GATCTGTACTGCTTAAACC 191

RESULT 2
AV692067/c 423 bp mRNA linear EST 16-JAN-2002

DEFINITION AV692067 GKC Homo sapiens cDNA clone GKCBLD04 5', mRNA sequence.
 ACCESSION AV692067
 VERSION AV692067.1 GI:102939930
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 423)
 AUTHORS Xu, X., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, J., Liu, F., Huang, Q., Cheng, Z., Li, N., Du, J., Hu, W., Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, J., Chen, Z., and Han, Z.
 TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
 MEDLINE 21625106
 PUBMED 11752456
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
 Tel: 86-21-50801919 (ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 FEATURES
 source
 1..423
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GKCBLD04"
 /tissue_type="hepatocellular carcinoma"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="GKC"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 1; Length 423;
 Best Local Similarity 90.9%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GATCTTGACTGCTTAACCC 22
 |||||
 133 GATGTTGACTGCTTAAGCC 112
 |||||
 RESULT 3
 LOCUS AU003892 505 bp mRNA linear EST 19-JAN-1999
 DEFINITION AU003892 Bombyx mori p50 (Daizo) Bombyx mori cDNA clone w800710, mRNA sequence.
 ACCESSION AU003892
 VERSION AU003892.1 GI:4161263
 KEYWORDS EST.
 SOURCE Bombyx mori (domestic silkworm)
 ORGANISM Bombyx mori
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditypsia; Bombycoidea; Bombycidae; Bombyx.
 REFERENCE 1 (bases 1 to 505)
 AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
 TITLE Establishment of cDNA database of Bombyx mori
 JOURNAL Unpublished (1999)
 COMMENT Contact: Mita K
 Genome Research Group
 National Institute of Agrobiological Sciences
 Owaehi 1-2, Tsukuba, Ibaraki 305-8634, Japan
 Email: kmita@nias.affrc.go.jp
 PROJECT = 'CREST project by JST'.

FEATURES
 source
 Location/Qualifiers
 1..505
 /organism="Bombyx mori"
 /mol_type="mRNA"
 /strain="p50 (Daizo)"
 /db_xref="taxon:7091"
 /clone="w800710"
 /clone_lib="Bombyx mori p50 (Daizo)"
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 1; Length 505;
 Best Local Similarity 90.5%; Pred. No. 6.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GATCTTGACTGCTTAACC 21
 |||||
 5 GTTCTTGACTGCTTAACC 25
 |||||
 RESULT 4
 LOCUS BZ419788 606 bp DNA linear GSS 10-DEC-2002
 DEFINITION BZ419788 1f59f06.b1 WGS-ZmayrF (DH5a methyl filtered) Zea mays genomic clone
 ACCESSION 1f59f06.5', genomic survey sequence.
 VERSION BZ419788
 KEYWORDS BZ419788.1 GI:26365470
 GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 606)
 AUTHORS Rabinowitz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N., Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L., Zutavern, T., McCombie, W.R. and Martienssen, R.A.
 TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)
 JOURNAL Unpublished (2002)
 COMMENT Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874
 Email: mcombie@cshl.org
 Plate: 1f59 row: f column: 06
 Seq primer: -21M13UnivFwd
 Class: Shotgun
 High quality sequence stop: 606.
 FEATURES
 source
 Location/Qualifiers
 1..606
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="1f59f06"
 /lab_host="DH5a"
 /clone_lib="WGS-ZmayrF (DH5a methyl filtered)"
 /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I; The vector was digested with Xba I and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19; b/g reads in pUC19). The same ligation was transformed into DH5a."
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 8; Length 606;
 Best Local Similarity 90.5%; Pred. No. 6.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 ATCTTGACTGCTTAACCC 22

```

Db      15 ATCTGTCTGCTTTAAACC 35
|||||
RESULT 5
LOCUS   AG402501
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-237J13.T7, genomic survey
sequence.
ACCESSION AG402501
VERSION   AG402501.1
KEYWORDS GSS.
SOURCE    Mus musculus molossinus
ORGANISM  Mus musculus molossinus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS  Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE    BAC end sequences of library MSMg01
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 711)
AUTHORS  Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suenho-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunya Abe (abe@rc.riken.jp).
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyada1, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
1..711
location/Qualifiers
1..711
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-237J13.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 80.9%; Score 17.8; DB 9; Length 711;
Best Local Similarity 90.5%; Pred. No. 6.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTTAAACC 22
|||||
Db 350 ATCTGTACTGCTTTAAACC 370

RESULT 6
LOCUS   CC638234
DEFINITION CGVBT41TH ZM 0.7.1.5 KB zea mays genomic clone ZMMBWA0498G10,
genomic survey sequence.
ACCESSION CC638234
VERSION   CC638234.1
KEYWORDS GSS.
SOURCE    Zea mays
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS  WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nuneberg,A., Robbins,D. and Lakey,N.
TITLE    Consortium for Maize Genomics
JOURNAL  Unpublished (2002)
COMMENT  Other_GSSs: OGVBT41TV
Contact: Cathy WhiteIaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whiteIaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
source
1..799
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBWA0498G10"
/clone_lib="ZM 0.7.1.5 KB"
/notes="vector: pBCK-; Site_1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 80.9%; Score 17.8; DB 9; Length 799;
Best Local Similarity 90.5%; Pred. No. 6.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTTAAACC 22
|||||
Db 67 ATCTGTCTGCTTTAAACC 87

RESULT 7
LOCUS   BG299973
DEFINITION HVSMA0022M13f Hordeum vulgare seedling shoot EST library
HVCNDA0001 (Cold stress) Hordeum vulgare subsp. vulgare cDNA clone
HVSMA0022M13f, mRNA sequence.
ACCESSION BG299973
VERSION   BG299973
KEYWORDS EST.
SOURCE    Hordeum vulgare subsp. vulgare
ORGANISM  Hordeum vulgare subsp. vulgare
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Pooidae; Triticeae; Hordeum.
1 (bases 1 to 804)
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
Yu,Y., Henry,D., Palmer,M., Rambo,T., Simons,J., Oates,R.,
Choi,D.W., Fenton,R.D. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex cold-stressed seedling shoot cDNA
library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 228
Seq primer: ATTTAACCTCACTAAGGG
High quality sequence start: 7
High quality sequence stop: 464.
Location/Qualifiers
1..804

```

```

/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEa0022M13f"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCMDNA0001 (Cold stress)"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under aseptic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 50C for 2 days. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 600000 plu
were in vivo excised to give Bluescript SK(-) cDNA
phagemide. These steps were performed in the TU close
laboratory at the University of California, Riverside
(Choi, Close, Fenton). Phagemids were plated and picked at
the Clemson University Genomics Institute (CUGI) (Begum,
Palmer, Friesch, Atkins and Wang). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Friesch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinbols A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9pages/bgr/31/cover.html)"

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 4; Length 804;
 Best Local Similarity 90.5%; Pred. No. 6.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGTACTGCTTAACC 21
 |||||
 Db 483 GTCTTGTACTGCTGAACC 503

RESULT 8

CG320249/c 895 bp DNA linear GSS 26-AUG-2003
 LOCUS CG3BP48TV_ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0770G24,
 DEFINITION genomic survey sequence.
 ACCESSION CG320249
 VERSION CG320249.1 GI:34237515
 KEYWORDS GSS.

SOURCE

Zea mays
 Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 895)

REFERENCE

Whiteclaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T.,
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
 Consortium for Maize Genomics
 Unpublished (2002)
 Other GSSs: CG3BP48TH
 Contact: Cathy Whiteclaw

COMMENT

7912 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whiteclaw@tigr.org

Seq primer: TF
 Class: sheared ends.
 Location/Qualifiers
 1..895

```

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0770G24"
/clone_lib="ZM_0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 895;
 Best Local Similarity 90.5%; Pred. No. 6.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAACC 22
 |||||
 Db 233 ATCTGTCTGTCTTAACC 213

RESULT 9
 CG234260 932 bp DNA linear GSS 22-AUG-2003
 LOCUS CG2A001TV_ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0656B01,
 DEFINITION genomic survey sequence.

ACCESSION CG234260
 VERSION CG234260.1 GI:34134146
 KEYWORDS GSS.

SOURCE

Zea mays
 Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 932)

REFERENCE

Whiteclaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T.,
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
 Consortium for Maize Genomics
 Unpublished (2002)
 Other GSSs: CG2A001TH
 Contact: Cathy Whiteclaw

COMMENT

7912 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whiteclaw@tigr.org
 Seq primer: TF
 Class: sheared ends.

FEATURES
 Location/Qualifiers
 1..932

```

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0656B01"
/clone_lib="ZM_0.7_1.5_KB"
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 932;
 Best Local Similarity 90.5%; Pred. No. 6.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGTACTGCTTAACC 22
 |||||
 Db 276 ATCTGTCTGTCTTAACC 296

RESULT 10

BP185852/c
LOCUS BP185852 500 bp mRNA linear EST 17-JUN-2003
DEFINITION BP185852 planarian head cDNA Dugesia japonica cDNA clone 03031_HH,
mRNA sequence.
ACCESSION BP185852
VERSION BP185852.1 GI:32899764
KEYWORDS EST.
SOURCE Dugesia japonica
ORGANISM Dugesia japonica
Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Sericata;
Tricladida; Paludicola; Dugesidae; Dugesia.
1 (bases 1 to 500)
Mineta, K., Nakazawa, M., Gebria, F., Ikeo, K., Agata, K. and
Gotohori, T.
Origin and evolutionary process of the CNS elucidated by
comparative genomics analysis of planarian ESTs
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7666-7671 (2003)
22709146
MEDLINE
PUBMED 12802012
COMMENT Contact: Katsuniko Mineta
National Institute of Genetics, Center for Information Biology and
DNA Data Bank of Japan
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6847
Fax: 81-559-81-6848
Email: kmineta@ab.nig.ac.jp
These clones and additional information are obtained from our web
site: <http://www.cdb.nig.ac.jp/dda/>.
Location/Qualifiers
1..500
/organism="Dugesia japonica"
/mol_type="mRNA"
/db_xref="taxon:6161"
/clone="03031_HH"
/issue_type="head"
/dev_stage="adult"
/clone_lib="planarian head cDNA"

ORIGIN
Query Match 79.1%; Score 17.4; DB 5; Length 500;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTTAA 19
|||||
DB 429 GATCTTGACTGCTTCAA 411

RESULT 11
CE153858/c
LOCUS CE153858 685 bp DNA linear GSS 25-SEP-2003
DEFINITION tigr-gss-dog-17000371358796 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION CE153858
VERSION CE153858.1 GI:35276514
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 685)
Kirkness, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
MEDLINE
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-638-0200

Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1..685
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BcXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 79.1%; Score 17.4; DB 9; Length 685;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTTGACTGCTTTAA 19
|||||
DB 425 GATCTTGACTGCTTTAA 407

RESULT 12
CG922134/c
LOCUS CG922134 700 bp DNA linear GSS 12-DEC-2003
DEFINITION MBET139TF mch2 Medicago truncatula genomic clone 60H5, genomic
survey sequence.
ACCESSION CG922134
VERSION CG922134.1 GI:39781812
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 700)
Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.
Sequencing of BAC ends from Medicago truncatula
Unpublished (2003)
Other GSSs: MBET139TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
Seq primer: TGTAAACGACGCGCAGT
Class: BAC ends.
Location/Qualifiers
1..700
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultiivar="genotype AL17"
/db_xref="taxon:3880"
/clone="60H5"
/clone_lib="mch2"
/note="Vector: pBelobAC11; Site 1: HindIII; Site 2:
HindIII; Cook, D.R. and Kim, D.U., unpublished"

ORIGIN
Query Match 79.1%; Score 17.4; DB 9; Length 700;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ATCTTGACTGCTTTAA 20
|||||
DB 173 ATCTTGACTGCTTCAAC 155

RESULT 13
AG452969

LOCUS AG452989 740 bp DNA linear GSS 04-JUN-2004
 DEFINITION Mus musculus molossinus DNA, clone:MSMg01-338H12.T7, genomic survey
 ACCESSION AG452989
 VERSION AG452989.1 GI:48154503
 KEYWORDS GSS.
 SOURCE Mus musculus molossinus
 ORGANISM Mus musculus molossinus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 BAC and Sequences of Library MSMg01
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 740)
 AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
 REFERENCE Direct Submission
 JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
 TITLE and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 AUTHORS 1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 JOURNAL Tel:81-45-503-9111, Fax:81-45-503-9170 //hgp.gsc.riken.go.jp/,
 (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
 Clones are derived from the mouse BAC library MSMg01. For BAC
 library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
 Tsukuba Institute, Bio Resource Center,
 The Institute of Physical and Chemical Research (RIKEN) 3-1-1
 Koyadai, Tsukuba, 305-0074 Japan
 phone: 81-298-36-9189, fax: 81-298-36-9199
 e-mail: abe@rtc.riken.jp
 PRIMERS
 Sequencing : T7
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI
 Location/Qualifiers
 1..740
 /organism="Mus musculus molossinus"
 /mol_type="genomic DNA"
 /sub_species="molossinus"
 /db_xref="taxon:57486"
 /clone="MSMg01-338H12.T7"
 /sex="male"
 /tissue_type="mixture of kidney and spleen"
 /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
 Query Match 79.1%; Score 17.4; DB 9; Length 740;
 Best Local Similarity 94.7%; Pred. No. 1e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CTTGTACTGCTTAACCC 22
 |||||
 Db 299 CTTGTACTGCTTAACCC 317

RESULT 14
 CB261729 242 bp mRNA linear EST 06-NOV-2003
 LOCUS CB261729
 DEFINITION 89-B8866-008-015-001-pB12 MP12-ADIS-008 Arabidopsis thaliana cDNA
 clone MP12p76700115Q 5-PRIME, mRNA sequence.
 ACCESSION CB261729
 VERSION CB261729.1 GI:32886502
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 242)
 REFERENCE 1 Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T.,
 Mitchell-Olds,T. and Weisshaar,B.
 TITLE Large-scale identification and analysis of genome-wide

JOURNAL single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
 MEDLINE Genome Res. 13 (6), 1250-1257 (2003)
 PUBMED 22683290
 COMMENT 12799357
 Contact: Weisshaar B
 ADIS DNA core facility at MP12
 Max-Planck-Institute for Plant Breeding Research
 Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Fax: 00492215062851
 Email: weisshaar@mp12-koeln.mpg.de
 Insert Length: 242 Std. Error: 0.00
 Plate: 15 row: 0 column: 01
 Seq primer: pb12; GTGGCGCGCCCTCTAG.
 Location/Qualifiers
 1..242
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /ecotype="C24"
 /db_xref="GABI:601306"
 /db_xref="taxon:3702"
 /clone="MP12p76700115Q"
 /tissue_type="seedling"
 /dev_stage="few days old seedlings"
 /lab_host="E. coli XL1-Blue MRF"
 /clone_lib="MP12-ADIS-008"
 /note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
 XhoI; cDNA library from Arabidopsis thaliana, accession
 C24; seedling; Lambda ZAPri phage library was made at the
 Max-Planck-Institute of Molecular Plant Physiology, Golm,
 Germany and mass-excised at the Max-Planck-Institute for
 Plant Breeding Research, Cologne, Germany; cloning sites
 EcoRI-XhoI; Note: Sequencing granted in the context of the
 GABI Arabidopsis Verbund I; Genetic Diversity,
 'Establishment of high-efficiency SNP-based mapping tools
 and development of methods for genome-wide mutation
 detection' PI: Bernd Weisshaar Sequence submission managed
 by RZPD/GABI-Primary databases: http://gabi.rzpd.de This
 clone is available from RZPD, contact RZPD (clone@rzpd.de)
 for further information."

ORIGIN
 Query Match 78.2%; Score 17.2; DB 6; Length 242;
 Best Local Similarity 86.4%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTTACTGCTTAACCC 22
 |||||
 Db 166 GATCTTACTGCTTAACCC 145

RESULT 15
 AU099200/c 300 bp mRNA linear EST 23-JAN-2004
 LOCUS AU099200
 DEFINITION Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
 HRC01728 similar to Human homeobox-containing protein mRNA, mRNA
 sequence.
 ACCESSION AU099200
 VERSION AU099200.1 GI:13550329
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 Suzuki,Y., Tsunoda,T., Sese,J., Taiza,H., Mizushima-Sugano,J.,
 Hata,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Suyama,A.,
 Sakaki,Y., Morishita,S., Okubo,K. and Sugano,S.
 Identification and characterization of the potential promoter
 regions of 1031 kinds of human genes
 Genome Res. 11 (5), 677-684 (2001)
 JOURNAL 21235182
 MEDLINE 21235182
 PUBMED 11337467
 COMMENT Contact: Yutaka Suzuki

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: y Suzuki@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and
Sugano, S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
149-156 (1997)

FEATURES

source

Location/Qualifiers

1..300

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HRC01728"

/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match

78.2%; Score 17.2; DB 1; Length 300;

Best Local Similarity 86.4%; Pred. No. 1.2e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

1 GATCTGTGACTGCTTAACCC 22

DB

165 GATCTGTGCTGCTTAACCTC 144

Search completed: February 8, 2005, 22:18:05
Job time : 270.041 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 423.751 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76
Sequence: 1 gattctgttcgctgcctccccc.....ccctgcataataatcaattc 76

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pac: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sbs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.2	72.6	5562	2	AB091787 Ocolemur.
2	54.2	71.3	212886	9	AC093946 Rattus no
3	54.2	71.3	234471	2	AC121424 Rattus no
4	52.6	69.2	363	11	BV089294 RPAMMSBO
5	52.6	69.2	363	11	BV097602 RPAMMSBO
6	52.6	69.2	463	10	DB3064 Mus musculu
7	52.6	69.2	695	11	BV089295 RPAMMSBO
8	52.6	69.2	695	11	BV097603 RPAMMSBO
9	52.6	69.2	9384	10	AF294397 Mus muscu
10	52.6	69.2	95826	10	AL805974 Mouse DNA
11	52	68.4	5684	9	AB091785 Lemur cat
12	52	68.4	8810	9	AB091786 Lemur cat
13	52	68.4	8810	9	AY040206 Homo sapi
14	52	68.4	158142	2	AL357130 Homo sapi
15	52	68.4	259202	9	AC002366 Human Xp2
16	51	67.1	6465	9	AB091781 Pan trogl
17	49.4	65.0	6451	4	AB091789 Bos tauru
18	49	64.5	6442	9	AB091783 Saimiri s
19	47.8	62.9	7425	4	AB091791 Sus scrofa

20	47.2	62.1	5712	4	AB091793 Equus cab
21	45.6	60.0	7163	9	AB091782 Pan trogl
22	45.6	60.0	38765	9	BS000568 Pan trogl
23	45.6	60.0	177654	2	AP000918 Homo sapi
24	45.6	60.0	177726	9	AC145770 Pan trogl
25	45.6	60.0	190089	9	BS000576 Pan trogl
26	45.6	60.0	200214	9	AC013412 Homo sapi
27	44.2	58.2	5591	4	AB091794 Equus cab
28	44.2	58.2	5151	9	AB091788 Equus cab
29	41.8	55.0	296	10	S74899 Mus musculu
30	41.8	55.0	727	10	MUSMAMEB Mus musculu
31	41.8	55.0	765	4	AB032194 Equus cab
32	41.8	55.0	789	4	AB032193 Equus cab
33	41.8	55.0	798	10	BC059090 Mus muscu
34	41.8	55.0	799	10	MUSMAMEA Mus musculu
35	41.8	55.0	6931	4	AB091792 Sus scrofa
36	41	53.9	6264	4	AB091790 Bos tauru
37	40.6	53.4	7454	9	AB091784 Saimiri s
38	40.2	52.9	65	6	CO531502 Sequence
39	40.2	52.9	399	10	RU070754 Rattus norv
40	40.2	52.9	457	10	RU060564 Rattus norv
41	40.2	52.9	476	6	AR452534 Sequence
42	40.2	52.9	753	10	RU060562 Rattus norv
43	40.2	52.9	780	10	U01245 Rattus norv
44	40.2	52.9	812	10	RU067130 Rattus norv
45	40.2	52.9	825	10	RU051195 Rattus norv

ALIGNMENTS

RESULT 1
AB091787
LOCUS
Ocolemur garnettii AMELX gene for amelogenin, partial cds.
ACCESSION
AB091787.1 GI:29126027
VERSION
KEYWORDS
SOURCE
ORGANISM
Ocolemur garnettii (small-eared galago)
Ocolemur garnettii
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsithini; Galagonidae; Ocolemur.

REFERENCE
1
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
JOURNAL
MEDLINE
PUBMED
22608569
12672962

REFERENCE
2 (bases 1 to 5562)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan kokusaimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@minsk.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)
FEATURES
source
1..5562
Location/Qualifiers
/organism="Ocolemur garnettii"
/mol_type="genomic DNA"
/db_xref="taxon:30611"
/sex="male"
/gene="AMELX"
/join(267..321,1401..1412)
/gene="AMELX"
/join(1413..1466,3385..3432,4699..4740,4832..4876,5125..5562)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66107.1"

Query Match 71.3%; Score 54.2; DB 2; Length 212866;
 Best Local Similarity 94.9%; Pred. No. 1.1e-09;
 Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GATTGTTGCTGCTGCTGAGACAGCCTTGTGCTATGCCCGTAGTAATAATACCCC 59
 112447 GATCTTGTGCTGCTGCTGAGACAGCTTTTGTGCTATGCCCGTAGTAATAACACCCC 112389

RESULT 3
 AC121424 234471 bp DNA 1linear HTG 21-SEP-2002
 LOCUS Rattus norvegicus clone CH230-122F17, *** SEQUENCING IN PROGRESS
 DEFINITION *** 2 unordered pieces.
 AC121424
 AC121424.3 'G1:23265350
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 234471)
 Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biwale, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Frazer, C. M., Gabiela, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregergys, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hu, Y., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Ledow, H., Levay, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshuber, L., Louised, H., Lozano, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Nandiasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwackemele, O., Okunolu, G., Olamunsgoon, A., Pal, S., Parks, K.,
 Pasernak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polinder, A., Popovic, D., Primus, E., Pu, L., L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shloman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Soza, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanu, K.,
 Vales, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, O., Wang, S., Warren, J., Warren, J., Wei, X., White, P.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, V., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G., and Gibbs, R. A.
 Direct Submission

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Unpublished
 2 (bases 1 to 234471)
 Worley, K. C.
 Direct Submission
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 234471)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

On Sep 21, 2002 this sequence version replaced gi:21908624.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXUI
 Center clone name: CH230-122F17
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 220465 bases at least Q40
 Consensus quality: 222968 bases at least Q30
 Consensus quality: 224416 bases at least Q20
 Estimated insert size: 241282; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 232989: contig of 232989 bp in length
 * 232990 233089: gap of unknown length
 * 233090 234471: contig of 1382 bp in length.
 Location/Qualifiers
 1. 234471
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10115"
 /clone="CH230-122F17"
 104912..105798
 /note="clone boundary
 clone end: T7
 site: EcoRI
 end_sequence: BH268493"
 complement(231872..232742)
 /note="clone boundary
 clone end: Sp6
 site: EcoRI
 end_sequence: BH268494"

ORIGIN
 Query Match 71.3%; Score 54.2; DB 2; Length 234471;
 Best Local Similarity 94.9%; Pred. No. 1.1e-09;
 Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 GATTGTTGCTGCTGCTGAGACAGCCTTGTGCTATGCCCGTAGTAATAATACCCC 59

exon 222..287
/product="amelogenin"
/number=2
286..463
/number=2

intron

ORIGIN

Query Match 69.2%; Score 52.6; DB 10; Length 463;
Best Local Similarity 93.2%; Pred. No. 1.5e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 59
|||||
Db 245 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 303
|||||

RESULT 7
BV089295/c 695 bp DNA linear STS 15-OCT-2003
LOCUS RPAMSE0001191 Roche Palo Alto Mus musculus STS genomic, sequence
DEFINITION
tagged site.
ACCESSION
BV089295 GI:37666774
VERSION
SV089295.1
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 695)
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
MUS musculus SNPs
TITLE
Unpublished (2003)
JOURNAL
COMMENT

CONTACT: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted
Primer B: No primer submitted.
Location/Qualifiers
1..695
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73613-72919-AL805974.7.1.95826"
/clone.lib="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains-of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D-H2/OSuJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MRL/MPJ, NZB/BinJ, NZW/Lac, SPRET/Ei.."
<1..>695

ORIGIN

STS

Query Match 69.2%; Score 52.6; DB 11; Length 695;
Best Local Similarity 93.2%; Pred. No. 1.6e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 59
|||||
Db 479 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 421
|||||

RESULT 9
AF294397 9384 bp DNA linear ROD 03-SEP-2000
LOCUS Mus musculus amelogenin gene, promoter and partial cds.
DEFINITION
ACCESSION
AF294397
VERSION
AF294397.1 GI:9965405
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 9384)
Snead,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
Paine,C.T., Chen,L.S., Burretein,J.M., Jitpukeedubudinta,S.,
White,S.N. and Bringsas,P. Jr.
Transgene animal model for protein expression and accumulation into
forming enamel
Connect. Tissue Res. 38 (1-4), 279-286 (1998)

JOURNAL
MEDLINE
20515040
PubMed
11063035
REFERENCE
2 (bases 1 to 9384)
Snead,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
Direct Submissi
Submitted (08-AUG-2000) Dentistry, University of Southern
California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
LOCATION/Qualifiers
1..9384
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

FEATURES
source

KEYWORDS
STS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 695)
Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
McPherson,J.D., Foernzler,D. and Peltz,G.
MUS musculus SNPs
TITLE
Unpublished (2003)
JOURNAL
COMMENT

CONTACT: Jonathan Usuka
Roche Palo Alto Genetics and Genomics Department
Roche Palo Alto
3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
Tel: 6508555807
Email: Jonathan.Usuka@roche.com
Primer A: No primer submitted with this STS
Primer B: No primer submitted with this STS.
Location/Qualifiers
1..695
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/map="X-73613-72919-AL805974.7.1.95826"
/clone.lib="Roche Palo Alto"
/note="SNPs developed from assay sequences derived from 15
different strains of mice (as of October 1, 2003). Those
strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D-H2/OSuJ,
BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
MRL/MPJ, NZB/BinJ, NZW/Lac, SPRET/Ei.."
<1..>695

ORIGIN

STS

Query Match 69.2%; Score 52.6; DB 11; Length 695;
Best Local Similarity 93.2%; Pred. No. 1.6e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 59
|||||
Db 479 GATTGTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGTAAATACCCC 421
|||||

RESULT 9
AF294397 9384 bp DNA linear ROD 03-SEP-2000
LOCUS Mus musculus amelogenin gene, promoter and partial cds.
DEFINITION
ACCESSION
AF294397
VERSION
AF294397.1 GI:9965405
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 9384)
Snead,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
Paine,C.T., Chen,L.S., Burretein,J.M., Jitpukeedubudinta,S.,
White,S.N. and Bringsas,P. Jr.
Transgene animal model for protein expression and accumulation into
forming enamel
Connect. Tissue Res. 38 (1-4), 279-286 (1998)

JOURNAL
MEDLINE
20515040
PubMed
11063035
REFERENCE
2 (bases 1 to 9384)
Snead,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
Direct Submissi
Submitted (08-AUG-2000) Dentistry, University of Southern
California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
LOCATION/Qualifiers
1..9384
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

FEATURES
source

QY 1 GATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 60
 DB 1526 GATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 1585

RESULT 12
 AB091786 8004 bp DNA linear PRI 02-MAY-2003
 LOCUS Lemur catla AMELY gene for amelogenin, partial cds.
 DEFINITION AB091786
 ACCESSION AB091786.1 GI:29126025
 VERSION
 KEYWORDS
 SOURCE Lemur catla (ring-tailed lemur)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Strepsirrhini; Lemnidae; Lemur.

REFERENCE
 AUTHORS Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
 TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal
 boundary in diverse mammalian species
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
 MEDLINE 22608569
 PUBMED 12672962
 REFERENCES 2 (bases 1 to 8004)
 AUTHORS Iwase, M., Saita, Y. and Takahata, N.
 TITLE Direct Submission
 JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
 Advanced Studies (Sokenkai), Department of Biosystems Science;
 Shonan Kokusai-mura, Hayama, Kanagawa 240-0193, Japan
 (E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
 Fax: 81-468-58-1544)
 FEATURES
 source
 Location/Qualifiers
 1..8004
 /organism="Lemur catla"
 /mol_type="genomic DNA"
 /db_xref="taxon:9447"
 /sex="male"
 1382..6505
 /gene="AMELY"
 join(1382..1437,3040..3051)
 /gene="AMELY"
 join(3052..3105,5192..5239,6461..6505)
 /gene="AMELY"
 /codon_start=1
 /product="amelogenin"
 /protein_id="BAC66106.1"
 /db_xref="GI:29126026"
 /translation="MGTWILFACLLGAAPAMPLQPHRGHYINFSYVLTPLKMYOS
 MIRPP"

ORIGIN
 Query Match 68.4%; Score 52; DB 9; Length 8004;
 Best Local Similarity 91.7%; Pred. No. 4.5e-09;
 Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 60
 DB 3063 GATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 3122

RESULT 13
 AY040206 8810 bp DNA linear PRI 10-APR-2002
 LOCUS Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
 DEFINITION AY040206
 ACCESSION AY040206.1 GI:15028582
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 8810)

AUTHORS Hart, P.S., Hart, T.C., Simmer, J.P. and Wright, J.T.
 TITLE A nomenclature for X-linked amelogenesis imperfecta
 JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)
 MEDLINE 21920287
 PUBMED 11922868
 REFERENCES 2 (bases 1 to 8810)
 AUTHORS Hart, S., Hart, T.C. and Wright, J.T.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
 3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA
 FEATURES
 source
 Location/Qualifiers
 1..8810
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp22.3-p22.1"
 1..8791
 /gene="AMELX"
 1..1416
 /gene="AMELX"
 1386..1391
 /gene="AMELX"
 join(1417..1472,2781..2846,4783..4830,6240..6281,
 6552..6977,8611..8791)
 /gene="AMELX"
 /product="amelogenin precursor"
 1417..1472
 /gene="AMELX"
 /number=1
 2781..2846
 /gene="AMELX"
 /number=2
 join(2793..2846,4783..4830,6240..6281,6552..6977,
 8611..8616)
 /gene="AMELX"
 /codon_start=1
 /product="amelogenin precursor"
 /protein_id="AAK77213.1"
 /db_xref="GI:15028583"
 /translation="MGTWILFACLLGAAPAMPLQPHRGHYINFSYVLTPLKMYOS
 IRPPSYGVBERGMGLNHQITLVLSOGRPTHTLOHHIIPVPAOQVYIPQPPMP
 VPGQSMPTIOHNPRLPRPAOQVYIPQVQVQPHQPMQPPVNHQVLPPLPLP
 MPQPPLPMLPDLTLEAMPSTDKTKEEVD"
 2793..2840
 /gene="AMELX"
 4783..4830
 /gene="AMELX"
 /number=3
 6107..6148
 /gene="AMELX"
 /number=4
 6240..6281
 /gene="AMELX"
 /number=5
 6552..6977
 /gene="AMELX"
 /number=6
 8611..8791
 /gene="AMELX"
 /number=7
 8785..8791
 /gene="AMELX"

ORIGIN
 Query Match 68.4%; Score 52; DB 9; Length 8810;
 Best Local Similarity 91.7%; Pred. No. 4.5e-09;
 Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 60
 DB 2804 GATTTATTTGCTGCTGCTGAGACAGCTTTGCTATGCCGTGAGTAAATACCCCT 2863

```

RESULT 14
AL357130/c
DEFINITION Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.
ACCESSION AL357130
VERSION AL357130.3 GI:9863814
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS McLay, K.
TITLE Direct Submisison
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9214076.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: b169E15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 156466 bases at least Q30
Consensus quality: 157038 bases at least Q20
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality
coverage: 4.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preapproved.
*
* 1 68897: contig of 68897 bp in length
* 68898 68897: gap of 100 bp
* 122843 122842: contig of 53845 bp in length
* 122843 122942: gap of 100 bp
* 122943 129584: contig of 6642 bp in length
* 129585 129684: gap of 100 bp
* 129685 158142: contig of 28458 bp in length.
Location/Qualifiers
1. 158142
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-169E15"
/clone_1fb="RPC1-13.1"
1. 68897
/note="assembly_fragment: 01108
fragment_chain:1
clone_end:T7
vector_side:left"
68998. 122842
/note="assembly_fragment: 00101
fragment_chain:1"
122943. 129584
/note="assembly_fragment: 00361
fragment_chain:1"
129685. 158142

```

```

/Note="assembly:fragment:01285
fragment chain:1
clone end:SP6
vector_side:right"

```

Query Match	68.4%;	Score 52;	DB 2;	Length 158142;
Best Local Similarity	91.7%;	Pred. No. 7.9e-09;		
Matches 55;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY	1	GATTTCGTCCTCCCTCCGAGACAGCCTTGCATACCCCGTAGTAATACCCCT	60	
DB	44314	GATTTCGTCCTCCCTCCGAGACAGCCTTTCGATGCGCTGTGATAACACCCCT	44255	
RESULT 15				
LOCUS	AC002366			
DEFINITION	Human Xp22 BAC CT-285115 (from Caltech/Research Genetics), PAC			
	RPc11-27c22 (from Roswell Park Cancer Center), and Cosmid U35B5			
	(from Lawrence Livermore), complete sequence.			
ACCESSION	AC002366	U79549	U70036	
VERSION	AC002366.1	GI:2739349		
KEYWORDS	HTG.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1 (bases 1 to 259202)			
	Muzny, R.A., Ansari-Lari, M.A., Timms, K.M., Yu, W., Dugan, S., Lu, J.,			
	Shen, Y., Rowland, K., Liu, W., Perez, J., Ding, Y., Haywood, M.,			
	Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,			
	Worley, K., Chen, E., Forcum, J., Arenson, A.D., Chiu, M.W.,			
	Grubbs, R.A., Brundage, E., Di, W., Chinault, C., Nelson, D. and			
	Gibbs, R.A.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 259202)			
AUTHORS	Chiu, M.W.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor			
REFERENCE	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
AUTHORS	3 (bases 1 to 259202)			
TITLE	Chiu, M.W.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (02-JAN-1998) Molecular and Human Genetics, Baylor			
AUTHORS	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
TITLE	4 (bases 1 to 259202)			
JOURNAL	Chiu, M.W.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (11-JUN-1998) Molecular and Human Genetics, Baylor			
TITLE	College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
JOURNAL	On Jan 2, 1998 this sequence version replaced gi:2642176.			
REFERENCE	Sequencing is completed to a minimum standard of double strand			
AUTHORS	coverage with a minimum of 2 clones and 2 reads with no ambiguities			
TITLE	or 2 chemistries with a minimum of 2 clones and 3 reads with no			
JOURNAL	ambiguities; if the sequence quality does not meet this standard,			
REFERENCE	it will be indicated in the annotation.			
AUTHORS	The repeat regions shown were identified using RepeatMasker by			
TITLE	Adrian Smit.			
JOURNAL	Sequence similarities were identified using Powerblast by Jinghui			
REFERENCE	Zhang.			
AUTHORS	Exon/Intron boundaries of identified genes were chosen if there			
TITLE	were canonical splice junctions that maintained sequence continuity			
JOURNAL	across the splice junctions.			
REFERENCE	Location/Qualifiers			
AUTHORS	1..259202			
TITLE	/organism="Homo sapiens"			
JOURNAL	/mol_type="Genomic DNA"			
REFERENCE	/db_xref="taxon:9606"			
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
J				

Wed Feb 9 09:28:18 2005

us-10-754-437-6.rge

Page 10

Best Local	Similarity	91.7%	Prod. No.	8.7e-09					
Matches	55	Conservative	0	Mismatches	5	Indels	0	Gaps	0
Qy	<pre> 1 GATTTTGTTCCTGCCTCCTCGAGAGAGCCTTGTCTATGCCCGTAGTAATAATACCCCT 60 69884 GATTTTATTCCTGCCTCCTCGAGAGAGCCTTGTGCATCCTGTGTGATTAACAACCCCT 69943 </pre>								
Db									

Search completed: February 8, 2005, 16:36:54
Job time : 426.751 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 100.817 Seconds
(without alignment)
3957.220 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 76
Sequence: 1 gattctgttgcctgcctcc.....ccctgcataatcattcattc 76

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	68.4	259202	12	ADQ18492
2	40.4	53.2	270	12	ADM80835
3	40.4	53.2	549	12	ADM80826
4	40.4	53.2	623	12	ADM80827
5	40.2	52.9	65	6	ABN28389
6	40.2	52.9	476	3	AAZ50832
7	39.2	51.6	556	10	ADB59026
8	39.2	51.6	556	10	ADB53782
9	38.8	51.1	722	6	AA141111
10	38.8	51.1	752	6	AA141110
11	37.2	48.9	318	12	ADM80836
12	37.2	48.9	750	2	AAZ07018
13	37.2	48.9	793	2	AAZ07018
14	35.6	46.8	802	2	AAZ07019
15	35.6	46.8	852	12	ADQ22977
16	27.6	36.3	1181	6	AB158182
17	27.6	36.3	1181	6	AB158182
18	26.8	35.3	849	12	ADJ48146
19	26.6	35.0	4031	4	AB103527
20	26.6	35.0	4940	4	AB103510
21	26.6	35.0	8256	4	AB103526

22	26.6	35.0	198161	6	ABK83564	ABK83564 Human cDN
23	26.6	35.0	198161	12	ADQ17348	ADQ17348 Human sof
24	26	34.2	795	6	AB190186	AB190186 Human pol
25	26	34.2	930	10	ADK68445	ADK68445 Simian T-
26	26	34.2	1802	2	AAT05712	AAT05712 STLV-PH96
27	26	34.2	3850	2	AAT05713	AAT05713 PR969-PTL
28	25.8	33.9	4713	8	ABX34656	ABX34656 Human mdd
29	25.8	33.9	5430	5	AA85365	AA85365 DNA encod
30	25.6	33.7	936	11	ABD00916	ABD00916 Klebsiell
31	25.4	33.4	591	2	AA212126	AA212126 Neisseria
32	25.4	33.4	591	3	AA254407	AA254407 Neisseria
33	25.4	33.4	1137	2	AA253817	AA253817 Neisseria
34	25.4	33.4	2000	8	ADA73001	ADA73001 Rice gene
35	25.4	33.4	77287	9	AA58261	AA58261 Murine tu
36	25.4	33.4	167739	9	AA58261	AA58261 Murine tu
37	25.2	33.2	274	10	ABX84738	ABX84738 Corn ear-
38	25.2	33.2	322	10	ABX87256	ABX87256 Corn ear-
39	24.6	32.4	350	4	AA11877	AA11877 Human bre
40	24.6	32.4	351	4	AA120767	AA120767 Human bre
41	24.6	32.4	7466	4	AB109192	AB109192 Drosophila
42	24.6	32.4	10825	6	ABN80207	ABN80207 Human che
43	24.6	32.4	32172	4	AA103473	AA103473 Human rep
44	24.4	32.1	313	4	AA119256	AA119256 Human bre
45	24.4	32.1	1410	10	ADC77539	ADC77539 Human HKE

ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
XX	ADQ18492;	
XX	26-AUG-2004 (first entry)	
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; ds.	
XX	Homo sapiens.	
XX	10-JUN-2004.	
XX	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002US-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
XX	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
XX	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

CC	neurological, developmental, connective tissue and cell proliferative disorders including cancer, e.g. breast, prostate, ovarian, lung or colorectal cancer, obesity and Tangier disease.
XX	
SQ	Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Query Match	53.2%; Score 40.4; DB 12; Length 549;
Best Local Similarity	88.0%; Pred. No. 3.3e-05;
Matches 44;	Conservative 0; Mismatches 6; Indels 0; Gaps 0
OY	1 GATTTCGTTGGCTGCCTCCTCGAGACAGCTTGTATGCCCGTAGTA 50 47 GATTTCATTTGCTCTGCTCTCTCGAGACAGCTTGTGCATCTCTGTGCTTA 96
Db	
RESULT 4	
ID	ADM80827 standard; cDNA; 623 BP.
XX	
AC	ADM80827;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human CADECIM-14 encoding cDNA SEQ ID NO:56.
KW	human; cell adhesion and extracellular matrix protein; CADECIM;
KW	neuroprotective; cytoskeletal; anorectic; immune disorder;
KW	neurological disorder; developmental disorder;
KW	connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW	Tangier disease; gene; BB.
XX	
OS	Homo sapiens.
FH	
Key	Location/Qualifiers
CDS	36..563
FT	/tag= a
FT	/product= "CADECIM-14"
XX	
PV	WO2004015396-A2.
XX	
PD	19-FEB-2004.
XX	
PF	12-AUG-2003; 2003WO-US025418.
XX	
PR	13-AUG-2002; 2002US-0403781P.
PR	30-AUG-2002; 2002US-0407034P.
PR	13-SEP-2002; 2002US-0410566P.
PR	24-SEP-2002; 2002US-0413482P.
PR	25-SEP-2002; 2002US-0413890P.
PR	08-NOV-2002; 2002US-0424904P.
PR	13-NOV-2002; 2002US-0426222P.
XX	
PA	(INCY-) INCYTE CORP.
XX	
EI	Ellicott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,
EI	Beccha SD, Marquis JP, Swarnaker A, Chawla NK, Ramkumar J,
EI	Hafelia AA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
EI	Mang JT, Chien D, Yang YG,
DR	WPI; 2004-191795/18.
DR	P-PDB; ADM80785.
XX	
PT	New cell adhesion and extracellular matrix proteins, useful in
PT	diagnosing, treating and preventing immune, neurological, developmental,
PT	connective tissue and cell proliferative disorders including cancer.
XX	
PS	Claim 5; SEQ ID NO 56; 272pp; English.
XX	
CC	The present sequence encodes a human cell adhesion and extracellular
CC	matrix protein designated CADECIM. CADECIM sequences has neuroprotective,
CC	cytostatic and anorectic activities. The CADECIM polypeptides and
CC	polynucleotides are useful in diagnosing, treating and preventing immune,
CC	neurological, developmental, connective tissue and cell proliferative

CC	disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC	cancer, obesity and Tangier disease.
XX	
SQ	Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
	Query Match 53.2%; Score 40.4; DB 12; Length 623;
	Best Local Similarity 88.0%; Pred. No.3.4e-05;
	Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0
OY	1 GATTTCGTTTGCCTGCCCTCGTGAGACAGACTTTGTCTATGCCCGTAGATA 50 47 GATTTCATTTCCTCCTCCTCGTGAGACAGCTTTTGGCATGCTGTGCTTA 96
DB	
RESULT 5	
ID	ABN28389 standard; DNA; 65 BP.
XX	ABN28389;
AC	
XX	ABN28389;
DT	
XX	15-JUN-2002 (first entry)
DE	
XX	Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.
KW	Human; mouse; rat; splice transcript; detection; RNA transcript; splice variant; transcriptome; oligonucleotide library; ss.
XX	
OS	Rattus norvegicus.
XX	
PN	WO200210449-A2.
PD	
XX	07-FEB-2002.
PF	
XX	20-JUN-2001; 2001WO-IB001903.
PR	
XX	28-JUL-2000; 2000US-0221607P.
PR	
XX	02-MAY-2001; 2001US-0287724P.
PA	(COMP-) COMPUGEN INC.
PI	
XX	Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
DR	
XX	WP; 2002-257383/30.
PT	New oligonucleotide libraries comprising oligonucleotides which
PT	selectively hybridize to mRNAs transcribed from a transcription unit of a
PT	genome; useful for detecting tissue-, pathology-, and developmental-
PT	specific genes.
PS	
XX	Example 1; SEQ ID NO 1137; 47pp; English.
XX	
CC	The present invention describes oligonucleotide libraries for detecting
CC	messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC)transcriptome comprises messenger RNAs transcribed from multiple
CC	transcription units that populate a genome. The library comprises several
CC	oligonucleotides, each capable of hybridising selectively to a set of
CC	messenger RNAs transcribed from a given transcription unit of the genome,
CC	which encodes one or more messenger RNA splice variants. The
CC	oligonucleotide libraries are useful for detecting mRNAs from a
CC	biological sample, in expression profiling studies, in qualitatively or
CC	quantitatively characterising the corresponding transcriptome, and in
CC	detecting RNA transcripts and splice variants of human or animal
CC	transcriptomes. The libraries may also be used as specialised mini
CC	libraries to detect transcripts of a sub-transcriptome under a particular
CC	biological or pathological state, and so allowing the detection of tissue
CC	- and pathology-specific genes such as those genes only expressed in
CC	specific tissue under a specific pathological condition; to detect
CC	developmental specific genes; and to detect RNA transcripts and splice
CC	variants of a transcriptome of a patient suffering from a particular
CC	disorder. ABN27253 to ABN9589 represent oligonucleotide sequences from
CC	rats, humans and mice, which are used in the exemplification of the
CC	present invention. N.B. The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format

CC directly from wipo at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;
Query Match 52.9%; Score 40.2; DB 6; Length 65;
Best Local Similarity 93.3%; Pred. No. 1.9e-05;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 45
6 GATCTGTGTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 50
DB
RESULT 6
AAZ50832
ID AAZ50832 standard; DNA, 476 BP.
XX
AC AAZ50832;
XX
DT 31-MAY-2000 (first entry)
XX
DE Rat amelogenin gene (A4).
XX
KW Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;
KW chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;
KW extracellular matrix protein; tooth enamel; enamel mineralisation;
KW ameloblast; bone regeneration; composite cell construct; ds.
XX
OS Rattus sp.
XX
FH Key
FT exon
FT Location/Qualifiers
FT 1..36
FT /*tag= a
FT /partial
FT /number= 1
FT 37..101
FT /*tag= b
FT /number= 2
FT 48..317
FT /*tag= c
FT /product= "Rat amelogenin protein"
FT /transl_except= (pos:69..71, aa:Gly)
FT 48..95
FT /*tag= d
FT /*tag= e
FT 96..314
FT /*tag= e
FT /label= Mature_rat_amelogenin_protein
FT 102..149
FT /*tag= f
FT /number= 3
FT 150..191
FT /*tag= g
FT /number= 4
FT 192..236
FT /*tag= h
FT /number= 5
FT 237..311
FT /*tag= i
FT /number= 6d
FT /note= "Comprises of gene segments 6a, b, c and d"
FT 312..317
FT /*tag= j
FT /number= 7
FT /note= "Includes the stop codon"
XX
PN WO200006734-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US017342.
XX
PR 29-JUL-1999; 98US-0094489P.
XX

PA (NOUN) UNTV NORTHWESTERN.
XX
PI Vels A, Nebgen DR;
XX
DR WPI; 2000-205464/18.
XX
DR P-PDSB; AAY45074.
XX
PT Novel amelogenin polypeptides and polynucleotides, useful for enhancing
PT bone generation in mammals and synthesizing bone matrix or articular
PT surfaces at implant sites.
XX
PS Example 2; Fig 11B; 79pp; English.
XX
CC The present DNA sequence is the full-length rat amelogenin gene (A4),
CC comprising exons 1-7, including the exon segment 6d. It is derived from
CC the rat incisor odontoblast-pulp cDNA library. The splice variants of
CC this gene functions as an osteogenic or chondrogenic inducing molecule
CC (CIM), which is useful for enhancing bone or cartilage growth. It has
CC osteopathic activity. Amelogenin belongs to the family of extracellular
CC matrix proteins, in developing tooth enamel, that are produced by the
CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or
CC osteogenic inducing amelogenin molecules are useful to induce
CC differentiation of cells to the osteogenic and chondrogenic phenotypes
CC and can be used in a composite cell construct for bone and cartilage
CC regeneration. The polynucleotides can be employed to produce the
CC polypeptides by recombinant techniques
XX
SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;
Query Match 52.9%; Score 40.2; DB 3; Length 476;
Best Local Similarity 93.3%; Pred. No. 3.7e-05;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 45
59 GATCTGTGTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 103
DB

RESULT 7
ADB59026
ID ADB59026 standard; DNA, 556 BP.
XX
AC ADB59026;
XX
DT 04-DEC-2003 (first entry)
XX
DE Toxicity-related gene, SEQ ID 4052.
XX
KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; ds.
XX
OS Unidentified.
XX
PN WO2003064624-A2.
XX
PD 07-AUG-2003.
XX
PF 31-JAN-2003; 2003WO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
XX
DR WPI, 2003-689530/65.
XX
PT Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to


```

PR the compound.
PS
PS Claim 1; SEQ ID NO 4052; 1156bp; English.
CC
CC The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
Query Match 51.6%; Score 39.2; DB 10; Length 556;
Best Local Similarity 91.1%; Pred. No. 9,5e-05;
Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 1 GATTTCGTTGGCTGCTCTCGGAGCAGCCCTTGCTATGCCCGT 45
Db 33 GATCTTCGTTGGCTGCTCTCTCGGAGCAGCTTTGATATGCCCT 77
RESULT 8
ADBS3782
ID ADBS3782 standard; DNA; 556 BP.
XX
XX ADBS3782;
AC
AC
DT 04-DEC-2003 (first entry)
XX
XX DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
XX
XX KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
XX OS Rattus norvegicus.
XX
XX WO2003065993-A2.
XX
XX 14-AUG-2003.
XX
XX PD 04-FEB-2003; 2003WO-US003462.
XX
XX
PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-036534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 09-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.

```

XX (GENE-) GENE LOGIC INC.
 XX Mendrick D, Porter M, Johnson K, Higgs B, Caetle A, Orr M;
 PI Elashoff M;
 PI WPI: 2003-731472/69.
 XX
 XX Determining if a compound induces a toxic effect on a tissue or cell, for
 PT identifying hepatotoxic compounds, comprises comparing a gene expression
 PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
 PT mean values.
 XX
 XX Claim 44; SEQ ID NO 4324; 874bp; English.
 XX
 XX The present invention describes a method for determining whether a
 CC compound induces a toxic effect on a tissue or cell. The method comprises
 CC preparing a gene expression profile of a tissue or cell sample exposed to
 CC the compound, and comparing the gene expression profile to a database
 CC comprising data or information on the Tox mean and non-Tox mean value.
 CC The method is useful for predicting or identifying at least one toxic
 CC effect, particularly hepatotoxicity, of a test or unknown compound. The
 CC genes listed in the specification are useful as diagnostic or toxicity
 CC markers for the prediction or identification of the physiological state
 CC of tissue or cell sample that has been exposed to a compound, or to
 CC identify or predict the toxic effects of a compound or an agent. These
 CC may also be used as markers for monitoring toxicity progression or for
 CC drug screening. The present sequence represents a primary rat hepatocyte
 CC toxicity modelling related gene sequence from the present invention.
 XX
 XX Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
 SQ
 XX
 XX Query Match 51.6%; Score 39.2; DB 10; Length 556;
 XX Best Local Similarity 91.1%; Pred. No. 9.5e-05;
 XX Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 GATTGTGGTGGCTGCTCCTCTGGAGAGAGCCTTGTATGCCCGT 45
 Db 33 GATCTGTGGCCCTGCTCCTCTGGAGAGAGCCTTGTATGCCCGT 77
 XX
 XX
 XX RESULT 9
 XX AAL41111
 XX ID AAL41111 standard; DNA; 722 BP.
 XX
 XX AAL41111;
 XX
 XX 16-OCT-2002 (first entry)
 DT
 XX
 XX gAML related Y-chromosome DNA sequence.
 DS
 XX
 XX Goat embryo sexual identification technique; goat amelogenin gene; gAML;
 KW sex-specific; gene; ds; Y-chromosome.
 KM
 XX
 XX Capra hircus.
 OS
 XX
 XX Key Location/Qualifiers
 FH 35..658
 FT CDS /*tag= a
 FT /product= "Y-chromosome protein"
 FT
 XX TW454013-A.
 XX
 XX 11-SEP-2001.
 XX
 XX 10-NOV-1999; 99TW-00119616.
 XX
 XX 10-NOV-1999; 99TW-00119616.
 XX
 XX (CHEN/) CHEN C.
 XX (JANG/) JANG J.
 XX (WENG/) WENG T.
 XX (JENG/) JENG D.

```

XX PI Chen C, Jang J, Weng T, Jeng D;
XX DR WPI; 2002-442016/47.
XX DR P-PSDB; AAO22534.
XX PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX PT identification, comprises high sensitivity even using single white blood
XX PT cell or cleavage c.
XX PS Disclosure; Page 28; 35pp; Chinese.
XX CC The invention relates to a goat embryo sexual identification technique
XX CC with high efficiency, sensitivity and repeatability. This technique
XX CC involves separately cloning and sequencing the coding regions and the
XX CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
XX CC results indicate that there are sex-specific sequences in the fifth
XX CC intron of the gene. The major characteristics according to the present
XX CC invention include high sensitivity, applicable in sex identification even
XX CC only using a single white blood cell or a single cleavage cell of
XX CC blastula; high diagnostic efficiency, capable of identifying hundreds of
XX CC goat embryo in 3 hours; simple operation procedures without complicated
XX CC steps of DNA extraction and need no additional control group intron; and
XX CC can be applied on different species of goats. This polynucleotide
XX CC sequence represents a GAML related Y-chromosome DNA sequence of the
XX CC invention
XX SQ Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 51.1%; Score 38.8; DB 6; Length 722;
XX Best Local Similarity 95.2%; Pred. No. 0.00015;
XX Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GATTTGTTTGCTGCTCTGGAGAGAGCTTGTATGCC 42
XX 46 GATTTGTTTGCTGCTCTGGAGAGAGCTTGTATGCC 87
XX
XX Db
XX
XX RESULT 10
XX ID AAL41110 standard; DNA; 752 BP.
XX AC AAL41110;
XX DT 16-OCT-2002 (first entry)
XX DE GAML related X-chromosome DNA sequence.
XX KW Goat embryo sexual identification technique; goat amelogenin gene; GAML;
XX KW sex-specific; gene; ds; X-chromosome.
XX OS Capra hircus.
XX FH Key Location/Qualifiers
XX FT CDS 35..658
XX FT /*tag= a
XX FT /product= "X-chromosome protein"
XX
XX TW454013-A.
XX PD 11-SEP-2001.
XX PF 10-NOV-1999; 99TW-00119616.
XX PR 10-NOV-1999; 99TW-00119616.
XX
XX (CHEN/) CHEN C.
XX PA (JANG/) JANG J.
XX PA (WENG/) WENG T.
XX PA (JENG/) JENG D.
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX

```

```

DR DR WPI; 2002-442016/47.
XX DR P-PSDB; AAO22534.
XX PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX PT identification, comprises high sensitivity even using single white blood
XX PT cell or cleavage c.
XX PS Disclosure; Page 28; 35pp; Chinese.
XX CC The invention relates to a goat embryo sexual identification technique
XX CC with high efficiency, sensitivity and repeatability. This technique
XX CC involves separately cloning and sequencing the coding regions and the
XX CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
XX CC results indicate that there are sex-specific sequences in the fifth
XX CC intron of the gene. The major characteristics according to the present
XX CC invention include high sensitivity, applicable in sex identification even
XX CC only using a single white blood cell or a single cleavage cell of
XX CC blastula; high diagnostic efficiency, capable of identifying hundreds of
XX CC goat embryo in 3 hours; simple operation procedures without complicated
XX CC steps of DNA extraction and need no additional control group intron; and
XX CC can be applied on different species of goats. This polynucleotide
XX CC sequence represents a GAML related X-chromosome DNA sequence of the
XX CC invention
XX SQ Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 51.1%; Score 38.8; DB 6; Length 752;
XX Best Local Similarity 95.2%; Pred. No. 0.00015;
XX Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GATTTGTTTGCTGCTCTGGAGAGAGCTTGTATGCC 42
XX 46 GATTTGTTTGCTGCTCTGGAGAGAGCTTGTATGCC 87
XX
XX Db
XX
XX RESULT 11
XX ID ADM80836 standard; cDNA; 318 BP.
XX AC ADM80836;
XX DT 03-JUN-2004 (first entry)
XX DE Human CADECM-23 encoding cDNA SEQ ID NO:65.
XX KW human; cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangier disease; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 36..260
XX FT /*tag= a
XX FT /product= "CADECM-23"
XX
XX WO2004015396-A2.
XX PD 19-FEB-2004.
XX PF 12-AUG-2003; 2003WO-US025418.
XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424904P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX
XX (INCY-) INCYTE CORP.
XX

```

XX Elliott VS, Khare R, Emerling BM, Kabie AE, Tran UK, Jin P;
PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Rankumar J;
PI Hatella AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
PI Wang JT, Chien D, Yang YG;
XX
DR WPI, 2004-191795/18.
DR P-PsDB; ADM60794.
XX
PT New cell adhesion and extracellular matrix proteins, useful in
PT diagnosing, treating and preventing immune, neurological, developmental,
PT connective tissue and cell proliferative disorders including cancer.
XX
PS Claim 5; SEQ ID NO 65; 272bp; English.
XX
CC The present sequence encodes a human cell adhesion and extracellular
CC matrix protein designated CADBCM. CADBCM sequences has neuroprotective,
CC cyostatic and anorectic activities. The CADBCM polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing immune,
CC neurological, developmental, connective tissue and cell proliferative
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and Tangle disease.
XX
SQ Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

Query_Match 48.9%; Score 37.2; DB 12; Length 318;
Best Local Similarity 92.9%; Pred. No. 0.00047;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

Oy 1 GATTTGTTCCTGCCTCCTCGAGACGGCCTTGCTATGCC 42
Db 47 GATTATTTGCTCTCCTCTCTGGAGACGCTTTCCTCATGCC 88

RESULT 12
AAZ07020
ID AAZ07020 standard; DNA; 750 BP.
XX
AC AAZ07020;
XX
DT 15-NOV-1999 (first entry)
DE Amelogenin X nucleotide sequence.
XX
KW Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KW multiple allelic site; apolipoprotein E; apoE; coronary artery disease;
XX Alzheimer's disease; ds.
OS Unidentified.
XX
PN WO9940226-A2.
XX
PD 12-AUG-1999.
XX
PF 08-JAN-1999; 99WO-US000499.
XX
PR 04-FEB-1998; 98US-00018595.
XX
PA (PEKE) PERKIN-ELMER CORP.
XX
PI Litvak KJ, Goodsaad F;
XX
DR WPI, 1999-539985/45.
XX
PT 5' nuclease amplification assay using fluorescence-quencher probes for
PT determination of a genotype at multiple allelic sites.
XX
PS Disclosure; Fig 10; 95pp; English.
XX
CC The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype

a sample of genomic DNA at two or more different allelic sites.
Generating a fluorescence spectrum and signature for each genotype, which uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of unknown sequences can be determined. The assay was shown to be useful for determining apoe genotypes. The assay can be used as a diagnostic tool for assessing the risk for coronary artery disease and/or late-onset Alzheimer's disease. Using the 5' nuclease assay of the invention it is possible to determine a genotype at two or more allelic sites in a single reaction. This approach is much faster than previous approaches to genotyping genes having two or more allelic sites, such as the apolipoprotein B gene. A key advantage of the method for determining the genotype of a sample of DNA at multiple allelic sites is that it does not rely on 5' nuclease assay working with 100% efficiency to distinguish between substantially homologous sequences such as alleles. The present sequence represent the nucleotide sequence for amelogenin X, which is used in the exemplification of the present invention

Dc Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Sd Query Match 48.9%; Score 37.2; DB 2; Length 750;
Best Local Similarity 92.9%; Pred. No. 0.00062;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGGCTCGCCTTCCTGGAGACGACTTGTGCATGCC 42
Db 80 GATTTAATTGGCCCTGCTCCTTGAGAGCACTTTGGCATGCC 121

RESULT 13
AAZ07018
ID AAZ07018 standard; DNA; 793 BP.

XX AC AAZ07018;
DT 15-NOV-1999 (first entry)

DE Amelogenin X nucleotide sequence.

KW Amelogenin Y; genotype; 5' nuclease amplification;
KW multiple allelic site; apolipoprotein E; apoE; coronary artery disease;
KM Alzheimer's disease; ss.

OS Unidentified.

PN MO9940226-A2.

PD 12-AUG-1999.

PP 08-JAN-1999; 99WO-US000499.

PR 04-FEB-1998; 98US-00018595.

PA (PERKE) PERKIN-ELMER CORP.

P1 Litvak KJ, Goodsaaid F;
PI WPI; 1999-539985/45.

DR 5' nuclease amplification assay using fluorescence-quencher probes for determination of a genotype at multiple allelic sites.

XX Disclosure; Fig BA; 95pp; English.

The present invention describes first and second sets of fluorescer-quencher probes used simultaneously in a 5' nuclease assay to identify which members of a first or second set of substantially homologous sequences are present in a DNA sample. The method can be used to genotype a sample of genomic DNA at two or more different allelic sites.
Generating a fluorescence spectrum and signature for each genotype, which uniquely reflects the assay's inherent inefficiency for that genotype
given the particular conditions, probes and primers used, the genotype of unknown sequences can be determined. The assay was shown to be useful for

CC determining apoe genotypes. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for ameloogenin X, which is
CC used in the exemplification of the present invention

SO Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 48.9%; Score 37.2; DB 2; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.0063;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTGCTGCGAGCAGCTTTGCTATGCC 42
DB 80 GATTTTATTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121

RESULT 14

AAZ07019
ID AAZ07019 standard; DNA; 802 BP.

AC AAZ07019;

DT 15-NOV-1999 (first entry)

DE Ameloogenin Y nucleotide sequence.

XX Ameloogenin X; Ameloogenin Y; genotype; 5' nuclease amplification;

KW multiple allelic site; apolipoprotein E; apoe; coronary artery disease;

KM Alzheimer's disease; 89.

XX Unidentified.

OS WC9940226-A2.

PN 12-AUG-1999.

PD 08-JAN-1999; 99WO-US000499.

PF 04-FEB-1998; 98US-00018595.

PR (PEKE) PERKIN-ELMER CORP.

PA Llavak KJ, Goodaaid F;

PI WPI; 1999-539985/45.

DR 5' nuclease amplification assay using fluorescence-quencher probes for

PT determination of a genotype at multiple allelic sites.

PS Disclosure; Fig 8B; 95pp; English.

XX The present invention describes first and second sets of fluorescer-

CC quencher probes used simultaneously in a 5' nuclease assay to identify

CC which members of a first or second set of substantially homologous

CC sequences are present in a DNA sample. The method can be used to genotype

CC a sample of genomic DNA at two or more different allelic sites.

CC Generating a fluorescence spectrum and signature for each genotype, which

CC uniquely reflects the assay's inherent inefficiency for that genotype

CC given the particular conditions, probes and primers used, the genotype of

CC unknown sequences can be determined. The assay was shown to be useful for

CC determining apoe genotypes. The assay can be used as a diagnostic tool

CC for assessing the risk for coronary artery disease and/or late-onset

CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is

CC possible to determine a genotype at two or more allelic sites in a single

CC reaction. This approach is much faster than previous approaches to

CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for ameloogenin Y, which is
CC used in the exemplification of the present invention

SO Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;

Query Match 46.8%; Score 35.6; DB 2; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.0026;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTGCTGCGAGCAGCTTTGCTATGCC 42
DB 80 GATTTTGTTCCTGCTGCTGCGAGCAGCTTTGCTATGCC 121

RESULT 15

ADQ22977
ID ADQ22977 standard; DNA; 852 BP.

AC ADQ22977;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

KW ds.

OS Homo sapiens.

PN WO2004048938-A2.

PD 10-JUN-2004.

PF 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

PA (PROT-) PROTEIN DESIGN LABS INC.

PI Aziz N, Ginsburg WM, Zlotnik A;

DR WPI; 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.

PS Example 2; SEQ ID NO 5797; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual,

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC DNA of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

SO Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;

Query Match 46.8%; Score 35.6; DB 12; Length 852;

Best Local Similarity 90.5%; Pred. No. 0.0027;

```
Matches      38; Conservative      0; Mismatches      4; Indels      0; Gaps      0;
```

QY 1 GATTTGTTGCTGCTCTGGAGCAGCCTTGTATGCC 42

Db 80 GATTTCCTTGTGGAGCAGCTTTGCCATGCC 121

Search completed: February 8, 2005, 13:57:00
Job time : 104.817 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 20.5262 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 1 gattctgttcctgcctcc.....ccctgcataataatcaatt 76

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/CTUS COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.2	52.9	476	4 US-09-744-128-3	Sequence 3, Appl1
2	37.2	48.9	500	2 US-09-018-595B-3	Sequence 3, Appl1
3	37.2	48.9	500	3 US-09-324-709A-3	Sequence 3, Appl1
4	37.2	48.9	793	2 US-09-018-595B-1	Sequence 1, Appl1
5	37.2	48.9	793	3 US-09-324-709A-1	Sequence 1, Appl1
6	35.6	46.8	802	2 US-09-018-595B-2	Sequence 2, Appl1
7	35.6	46.8	802	3 US-09-324-709A-2	Sequence 2, Appl1
8	25.6	33.7	936	4 US-09-489-039A-6691	Sequence 6691, Ap
9	25.2	33.2	274	4 US-09-313-294A-3198	Sequence 3198, Ap
10	25.2	33.2	277	4 US-09-313-294A-5716	Sequence 5716, Ap
11	24.2	31.8	930	4 US-09-710-279-1607	Sequence 1607, Ap
12	24.2	31.8	1311	3 US-09-134-001C-386	Sequence 386, App
13	24.2	31.8	1315	4 US-09-710-279-3881	Sequence 3881, App
14	24	31.6	1107	4 US-09-252-991A-4254	Sequence 4254, Ap
15	23.6	31.1	576	4 US-09-489-039A-4185	Sequence 4185, Ap
16	23.4	30.8	684	4 US-09-270-767-7458	Sequence 7458, Ap
17	23.4	30.8	684	4 US-09-270-767-22740	Sequence 22740, A
18	23.4	30.8	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
19	23.4	30.8	1664976	4 US-09-692-570-1	Sequence 1, Appl1
20	23	30.3	1386	4 US-09-543-681A-2276	Sequence 2276, Ap
21	22.8	30.0	826	6 US-09-189-760-5	Patent No. 5252477
22	22.8	30.0	1529	3 US-09-188-811-5	Sequence 5, Appl1
23	22.8	30.0	1529	3 US-09-514-422-5	Sequence 5, Appl1
24	22.8	30.0	1529	3 US-09-514-422-5	Sequence 5, Appl1
25	22.8	29.7	2791	4 US-09-799-451-500	Sequence 500, App
26	22.6	29.7	461	4 US-09-513-999C-1419	Sequence 1419, Ap
27	22.6	29.7	1080	6 US-09-223-3	Patent No. 5260223

C 28	22.6	29.7	2039	4 US-09-799-451-583	Sequence 583, App
C 29	22.6	29.7	2087	4 US-09-799-451-582	Sequence 582, App
C 30	22.6	29.7	1830121	4 US-09-557-884-1	Sequence 1, Appl1
C 31	22.6	29.7	1830121	4 US-09-643-990A-1	Sequence 1, Appl1
C 32	22.6	29.7	1830121	4 US-10-329-960-1	Sequence 1, Appl1
C 33	22.4	29.5	561	4 US-09-248-796A-14014	Sequence 14014, A
C 34	22.4	29.5	681	4 US-09-134-000C-2467	Sequence 2467, Ap
C 35	22.4	29.5	3860	4 US-09-620-312D-873	Sequence 873, App
C 36	22.4	29.5	7195	3 US-08-478-507-6	Sequence 6, Appl1
C 37	22.4	29.5	7195	3 US-09-128-275A-6	Sequence 6, Appl1
C 38	22.4	29.5	7195	3 US-09-553-427-6	Sequence 6, Appl1
C 39	22.4	29.5	14636	3 US-09-173-914-6	Sequence 6, Appl1
C 40	22.4	29.5	786431	4 US-09-751-389-3	Sequence 3, Appl1
C 41	22.2	29.2	11933	3 US-09-470-618-13	Sequence 13, Appl1
C 42	22.2	29.2	11933	3 US-09-364-862-13	Sequence 13, Appl1
C 43	22.2	29.2	161652	4 US-09-497-855A-40	Sequence 40, Appl1
C 44	22.2	28.9	592	3 US-08-896-164-80	Sequence 80, Appl1
C 45	22.2	28.9	667	4 US-09-221-017B-597	Sequence 597, App

ALIGNMENTS

```

RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3
Query Match          52.9%; Score 40.2; DB 4; Length 476;
Best Local Similarity 93.3%; Pred. No. 2.6e-06;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGCTGAGCAGCCTTGTATGCCCGT 45
Db 59 GATCTGTGCTGCTGCTGCTGAGCAGCCTTGTATGCCCGT 103

RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Weitz,
; ADDRESSER: Wilson Somsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA

```

ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-018-595B-3

Query Match 48.9%; Score 37.2; DB 2; Length 500;
Best Local Similarity 92.9%; Pred. No. 4,1e-05;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCC 42
DB 80 GATTTATTGCTGCTGCTGAGAGAGCTTTGCTATGCC 121

RESULT 3
US-09-324-709A-3
Sequence 3, Application US/09324709A
Patent No. 6134707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
TITLE OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-324-709A-3

Query Match 48.9%; Score 37.2; DB 3; Length 500;
Best Local Similarity 92.9%; Pred. No. 4,1e-05;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCC 42
DB 80 GATTTATTGCTGCTGCTGAGAGAGCTTTGCTATGCC 121

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

Query Match 48.9%; Score 37.2; DB 2; Length 793;
Best Local Similarity 92.9%; Pred. No. 5e-05;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCC 42
DB 80 GATTTATTGCTGCTGCTGAGAGAGCTTTGCTATGCC 121

RESULT 5
US-09-324-709A-1
Sequence 1, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-1
Query Match 48.9%; Score 37.2; DB 3; Length 793;
Best Local Similarity 92.9%; Pred. No. 5e-05; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GATTTGTTGCTGCTCTGCTGGAGACAGCTTTGCTATGCC 42
Db 80 GATTTATTGCTGCTGCTCTGGAGACAGCTTTGCTATGCC 121
RESULT 6
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2
Query Match 46.8%; Score 35.6; DB 2; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.00021;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 GATTTGTTGCTGCTCTGCTGGAGACAGCTTTGCTATGCC 42
Db 80 GATTTGTTGCTGCTCTGCTGGAGACAGCTTTGCTATGCC 121
RESULT 7
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362

REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match 46.8%; Score 35.6; DB 3; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.00021;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGTTCCTGCTGCTGCTGAGACGCTTTGCTATGCC 42
DB 80 GATTTTGTTCCTGCTGCTGCTGAGACGCTTTGCTATGCC 121

RESULT 8
US-09-489-039A-6691/c
Sequence 6691, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 6691
LENGTH: 936
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6691

Query Match 33.7%; Score 25.6; DB 4; Length 936;
Best Local Similarity 62.5%; Pred. No. 2;
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 13 CTGCTCTCTGGAGACGCTTTGCTATGCCCGGTAAGTAATACCTTGATATATATCA 72
DB 100 CCGCCGCGCGGCGAGAGCTGTGGTATCAAGACCGCAATAGACCTGATGCGATCA 41

QY 73 ATTT 76
DB 40 GTTT 37

RESULT 9
US-09-313-294A-3198/c
Sequence 3198, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Itzudl, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 3198
LENGTH: 274
TYPE: DNA
ORGANISM: Zea mays
FEATURE:

NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700611080H1
US-09-313-294A-3198

Query Match 33.2%; Score 25.2; DB 4; Length 274;
Best Local Similarity 71.7%; Pred. No. 1.8;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 31 CTTTGCTATGCCCGGAGTAATAATACCTTGATATATTCATT 76
DB 61 CTTTGCTATGCCGTTGAGTAATAAATCCTGATGATTCATT 16

RESULT 10
US-09-313-294A-5716/c
Sequence 5716, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Itzudl, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 5716
LENGTH: 277
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6476212 700350567H1
NAME/KEY: unsure
LOCATION: 51
OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5716

Query Match 33.2%; Score 25.2; DB 4; Length 277;
Best Local Similarity 71.7%; Pred. No. 1.8;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 31 CTTTGCTATGCCCGGAGTAATAATACCTTGATATATTCATT 76
DB 246 CTTTGCTATGCCGTTGAGTAATAAATCCTGATGATTCATT 201

RESULT 11
US-09-710-279-1607
Sequence 1607, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P03480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1607
LENGTH: 930
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1607

Query Match 31.8%; Score 24.2; DB 4; Length 930;
Best Local Similarity 62.3%; Pred. No. 7.1;

Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGGAGTAAATACCTCGATATAT 69

Db 345 TGCATTTTTCATATGTCGACGCTTAATTAAGCCCGGAGTGCATTAACATTCATATATGT 404

QY 70 T 70

Db 405 T 405

RESULT 12
US-09-134-001C-386

; Sequence 386, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 386

; LENGTH: 1311

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-386

Query Match

Best Local Similarity 31.8%; Score 24.2; DB 3; Length 1311;

Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGGAGTAAATACCTCGATATAT 69

Db 726 TGCATTTTTCATATGTCGACGCTTAATTAAGCCCGGAGTGCATTAACATTCATATATGT 785

QY 70 T 70

Db 786 T 786

RESULT 13

US-09-710-279-3881/c

; Sequence 3881, Application US/09710279

; Patent No. 6703492

; GENERAL INFORMATION:

; APPLICANT: KIMBERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/09/710,279

; PRIOR FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 3881

; LENGTH: 3155

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

US-09-710-279-3881

QY 10 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGGAGTAAATACCTCGATATAT 69

Db 2712 TGCATTTTTCATATGTCGACGCTTAATTAAGCCCGGAGTGCATTAACATTCATATATGT 2653

QY 70 T 70

Db 2652 T 2652

RESULT 14
US-09-252-991A-4254/c

; Sequence 4254, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 4254

; LENGTH: 1107

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-4254

Query Match

Best Local Similarity 31.6%; Score 24; DB 4; Length 1107;

Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 6 TGGTTCCTGCTCTCGGAGCAGCCTTTGCTATGCCCGG 45

Db 1104 TTTTCTCTTCCCTCTTGAAGGCGCATTTGCTTTGCTGT 1065

RESULT 15

US-09-489-039A-4185/c

; Sequence 4185, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 4185

; LENGTH: 576

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-4185

Query Match

Best Local Similarity 31.1%; Score 23.6; DB 4; Length 576;

Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 14 TGGCTGCTCTCGGAGCAGCCTTTGCTATGCCCGGAGTAAATACCTCGATATATTTCA 73

Db 540 TGGTTCCTGCTCTCGGAGCAGCCTTTGCTATGCCCGGAGTAAATACCTCGATATATTTCA 481

QY 74 TT 75

Db 480 TT 479

Wed Feb 9 09:28:19 2005

us-10-754-437-6.rni

Page 6

Search completed: February 8, 2005, 22:26:03
Job time : 26.6373 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: February 8, 2005, 16:37:21 ; Search time 102.41 Seconds
(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-6

Perfect score: 1 gattctgttcgtcgtcctcc.....ccctgcataatcatcattt 76

Sequence: 1 gattctgttcgtcgtcctcc.....ccctgcataatcatcattt 76

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PC7_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PC7US_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52	68.4	600	US-10-027-632-287049	Sequence 287049,
C 2	52	68.4	600	US-10-027-632-287050	Sequence 287050,
C 3	52	68.4	600	US-10-027-632-287049	Sequence 287049,
C 4	52	68.4	600	US-10-027-632-287050	Sequence 287050,
5	52	68.4	259202	US-10-723-860-1311	Sequence 1311, Ap
6	40.2	52.9	65	US-09-908-975-1137	Sequence 1137, Ap
7	37.2	48.9	500	US-10-104-774-3	Sequence 3, Appli
8	37.2	48.9	500	US-10-104-774-3	Sequence 3, Appli
9	37.2	48.9	793	US-10-104-774-1	Sequence 1, Appli
10	37.2	48.9	793	US-10-104-774-1	Sequence 1, Appli
11	37.2	48.9	793	US-10-755-889-605	Sequence 605, App
12	35.6	46.8	802	US-10-104-774-2	Sequence 2, Appli

13	35.6	46.8	802	US-10-455-150-2	Sequence 2, Appli
14	35.6	46.8	852	US-10-723-860-5797	Sequence 5797, Ap
C 15	28.8	37.9	129	US-10-437-963-53858	Sequence 53858, A
C 16	28.2	37.1	225883	US-10-1-175-523-57	Sequence 57, Appl
C 17	27.6	36.3	1183	US-10-432-241-6	Sequence 6, Appli
C 18	27.6	36.3	1881	US-10-432-241-2	Sequence 2, Appli
19	26.8	35.3	849	US-10-389-566-150	Sequence 150, App
20	26.6	35.0	198161	US-10-775-169-52	Sequence 52, Appl
21	26.6	35.0	198161	US-10-723-860-165	Sequence 165, App
C 22	26.2	34.5	640	US-10-767-701-3691	Sequence 3691, App
C 23	26	34.2	795	US-10-264-237-748	Sequence 748, App
C 24	25.8	33.9	2415	US-10-425-115-114894	Sequence 114894, A
C 25	25.8	33.9	654	US-10-027-632-33959	Sequence 33959, A
C 26	25.6	33.7	654	US-10-027-632-33959	Sequence 33959, A
27	25.6	33.7	739	US-10-027-632-146671	Sequence 146671, A
28	25.6	33.7	739	US-10-027-632-146672	Sequence 146672, A
29	25.6	33.7	739	US-10-027-632-146671	Sequence 146671, A
30	25.6	33.7	739	US-10-027-632-146672	Sequence 146672, A
31	25.6	33.7	763	US-10-027-632-170284	Sequence 170284, A
32	25.6	33.7	763	US-10-027-632-170284	Sequence 170284, A
C 33	25.4	33.4	818	US-10-027-632-9912	Sequence 9912, Ap
C 34	25.4	33.4	818	US-10-027-632-9913	Sequence 9913, Ap
C 35	25.4	33.4	818	US-10-027-632-9913	Sequence 9913, Ap
C 36	25.4	33.4	818	US-10-027-632-9913	Sequence 9913, Ap
37	25.4	33.4	2025	US-10-437-963-73382	Sequence 73382, A
38	25.2	33.2	183	US-10-674-124A-4652	Sequence 4652, Ap
C 39	25.2	33.2	1339	US-10-425-114-31894	Sequence 31894, A
C 40	25.2	33.2	2154	US-10-425-115-37543	Sequence 37543, A
C 41	24.8	32.6	540	US-10-027-632-307914	Sequence 307914, A
C 42	24.8	32.6	307914	US-10-027-632-307914	Sequence 307914, A
C 43	24.6	32.4	405	US-10-198-846-3322	Sequence 3322, Ap
C 44	24.6	32.4	488	US-10-437-963-56587	Sequence 56587, A
C 45	24.6	32.4	676	US-10-027-632-237425	Sequence 237425, A

ALIGNMENTS

RESULT 1
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
Query Match 68.4%; Score 52; DB 13; Length 600;
Best Local Similarity 91.7%; Pred. No. 2.5e-10;

Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 60
|||||
Db 245 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 186

RESULT 2

US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:

APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050

Query Match 68.4%; Score 52; DB 13; Length 600;
Best Local Similarity 91.7%; Pred. No. 2.5e-10;
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 60
|||||
Db 245 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 186

RESULT 3

US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

Query Match

68.4%; Score 52; DB 15; Length 600;
Best Local Similarity 91.7%; Pred. No. 2.5e-10;
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 60
|||||
Db 245 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 186

RESULT 4

US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:

APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050

Query Match 68.4%; Score 52; DB 15; Length 600;
Best Local Similarity 91.7%; Pred. No. 2.5e-10;
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 60
|||||
Db 245 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGTAGTAATACCCCT 186

RESULT 5

US-10-723-860-1311
; Sequence 1311, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:

APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26

PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1311
LENGTH: 259202
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 68.4%; Score 52; DB 18; Length 259202;
Best Local Similarity 91.7%; Pred. No. 1.7e-09;
Matches 55; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTTGCTATGCCCTGTAATACCCCT 60
DB 69884 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCCCTGTAATACCCCT 69943

RESULT 6

US-09-908-975-1137
Sequence 1137, Application US/09908975
Publication No. US20030165843A1
GENERAL INFORMATION:
APPLICANT: SHOSHAN, Avi
APPLICANT: MASSERMAN, Alon
APPLICANT: MINTZ, Eli
APPLICANT: MINTZ, Eli
APPLICANT: FAIGLER, Simchon
TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
FILE REFERENCE: 36688-0005
CURRENT APPLICATION NUMBER: US/09/908,975
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US 60/287,724
PRIOR FILING DATE: 2001-05-02
PRIOR APPLICATION NUMBER: US 60/221,607
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 32337
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1137
LENGTH: 65
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-908-975-1137

Query Match 52.9%; Score 40.2; DB 10; Length 65;
Best Local Similarity 93.3%; Pred. No. 6.6e-06;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTTGCTATGCCCT 45
DB 6 GATCTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCCCT 50

RESULT 7

US-10-104-774-3
Sequence 3, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 500
TYPE: DNA

ORGANISM: Homo sapiens
US-10-104-774-3

Query Match 48.9%; Score 37.2; DB 13; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.0002;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121

RESULT 8

US-10-455-150-3
Sequence 3, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft windows 95
SOFTWARE: wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match 48.9%; Score 37.2; DB 16; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.0002;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121

RESULT 9

US-10-104-774-1
Sequence 1, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-1

Query Match 48.9%; Score 37.2; DB 13; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.00023;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTTATTGGCTGCTGCTGAGAGAGAGCTTTGCCATGCC 121

RESULT 10
US-10-455-150-1
Sequence 1, Application US/10455150
Publication No. US2004005302A1
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Weitz,
Wilson Soncini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:

NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match 48.9%; Score 37.2; DB 16; Length 793;

Best Local Similarity 92.9%; Pred. No. 0.00023;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTTATTGGCTGCTGCTGAGAGAGAGCTTTGCCATGCC 121

RESULT 11
US-10-755-889-605
Sequence 605, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF- κ B
PATHWAY

FILE REFERENCE: D0284 NP

CURRENT APPLICATION NUMBER: US/10/755,889

CURRENT FILING DATE: 2004-01-13

PRIOR APPLICATION NUMBER: U.S. 60/440,068

PRIOR FILING DATE: 2003-01-14

PRIOR APPLICATION NUMBER: U.S. 60/469,757

PRIOR FILING DATE: 2003-05-12

NUMBER OF SEQ ID NOS: 823

SOFTWARE: PatentIn version 3.2

SEQ ID NO 605

LENGTH: 793

TYPE: DNA

ORGANISM: Homo sapiens

US-10-755-889-605

Query Match 48.9%; Score 37.2; DB 17; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.00023;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTTATTGGCTGCTGCTGAGAGAGAGCTTTGCCATGCC 121

RESULT 12
US-10-104-774-2
Sequence 2, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division

TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE

FILE REFERENCE: 16842-782

CURRENT APPLICATION NUMBER: US/10/104,774

CURRENT FILING DATE: 2002-03-21

PRIOR APPLICATION NUMBER: US 09/018,595

PRIOR FILING DATE: 1998-02-04

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 802

TYPE: DNA

ORGANISM: Homo sapiens

US-10-104-774-2

Query Match 46.8%; Score 35.6; DB 13; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.001;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTTGTTGGCTGCTGCTGAGAGAGAGCTTTGCCATGCC 121

RESULT 13
US-10-455-150-2
Sequence 2, Application US/10455150


```
/ Publication No. US20040053302A1
/ GENERAL INFORMATION:
/ APPLICANT: Perkin-Elmer Corporation,
/ TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
/ SITES
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: David J. Weitz,
/ Wilson Sonsini Goodrich & Rosati
/ STREET: 650 Page Mill Road
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94304-1050
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5 inch diskette
/ COMPUTER: IBM compatible
/ OPERATING SYSTEM: Microsoft Windows 95
/ SOFTWARE: Wordperfect for windows 6.0,
/ ASCII (DOS) TEXT format
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/455,150
/ FILING DATE: 04-Jun-2003
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/324,709A
/ FILING DATE: 03-June-1999
/ ATTORNEY/AGENT INFORMATION:
/ NAME: David J. Weitz
/ REGISTRATION NUMBER: 38,362
/ REFERENCE/DOCKET NUMBER: 16842-758
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 493-9300
/ TELEFAX: (650) 493-6811
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 802 nucleotides
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-455-150-2

Query Match          46.8%; Score 35.6; DB 16; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.001;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGAGAGAGCCTTGCTATGCC 42
DB      80 GATTTGTTGCTGCTGCTGAGAGAGCCTTTGCAATGCC 121

RESULT 14
US-10-723-860-5797
/ Sequence 5797, Application US/10723860
/ Publication No. US20040253606A1
/ GENERAL INFORMATION:
/ APPLICANT: Aziz, Nacasha
/ APPLICANT: Ginsburg, Wendy M.
/ APPLICANT: Zlotnick, Albert
/ TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
/ FILE REFERENCE: 05882.0193.NPUS01
/ CURRENT APPLICATION NUMBER: US/10/723,860
/ PRIOR FILING DATE: 2003-11-26
/ PRIOR APPLICATION NUMBER: 60/429,739
/ PRIOR FILING DATE: 2002-11-26
/ NUMBER OF SEQ ID NOS: 8393
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 5797
/ LENGTH: 852
/ TYPE: DNA
```

```
/ ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match          46.8%; Score 35.6; DB 18; Length 852;
Best Local Similarity 90.5%; Pred. No. 0.001;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGAGAGAGCCTTGCTATGCC 42
DB      80 GATTTGTTGCTGCTGCTGAGAGAGCCTTTGCAATGCC 121

RESULT 15
US-10-437-963-53858/c
/ Sequence 53858, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovall, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 53858
/ LENGTH: 129
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_56017C.1
US-10-437-963-53858

Query Match          37.9%; Score 28.8; DB 17; Length 129;
Best Local Similarity 69.6%; Pred. No. 0.3;
Matches 39; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      3 TTTTGTGCTGCTGCTGCTGAGAGAGCCTTGCTATGCCCGTAGATAATGCC 58
DB      118 TATTTGCTGCTGCTGCTGAGAGAGGTTCTTCATGCCCAATGAATAAGAAC 63

Search completed: February 9, 2005, 06:50:58
Job time : 104.41 secs
```

This Page Blank (uspto)

ORIGIN

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Query Match 55.0%; Score 41.8; DB 7; Length 303;
Best Local Similarity 95.6%; Pred. No. 0.00046;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 45
Db 81 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 125

RESULT 2

W40649 430 bp mRNA linear EST 11-SEP-1996
LOCUS me43h12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:351335 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.

ACCESSION W40649 GI:1324982
VERSION W40649.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)

REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheielsenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:22135
Seq primer: ETPprimer
High quality sequence stop: 323.

FEATURES

source
Location/Qualifiers
1..430

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:351335"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCCATCTGAGTGGAGGAGCGGCGGCTTTTCTATGCCCT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match

55.0%; Score 41.8; DB 7; Length 430;

Best Local Similarity 95.6%; Pred. No. 0.0005;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 45
Db 69 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 113

RESULT 3

W36345 536 bp mRNA linear EST 11-SEP-1996
LOCUS mb22c12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.

ACCESSION W36345 GI:1318120
VERSION W36345.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 536)

REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Scheielsenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMT Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:216366
Seq primer: ETPprimer
High quality sequence stop: 359.

FEATURES

source

Location/Qualifiers
1..536
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:334966"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCCATCTGAGTGGAGGAGCGGCGGCTTTTCTATGCCCT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 55.0%; Score 41.8; DB 7; Length 536;
Best Local Similarity 95.6%; Pred. No. 0.00053;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 45
Db 64 GATTTGTTGGCTGCTGCTGAGAGAGCTTTGCTATGCCCT 108

RESULT 4

LOCUS	AY19441	585 bp	DNA	linear	GSS 17-DEC-2003
DEFINITION	Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,				
ACCESSION	AY19441				
VERSION	AY19441.1	GI:39775398			
KEYWORDS	GSS.				
SOURCE					
ORGANISM	Mus musculus (house mouse)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 585)				
TITLE	Clark,A.G., Gnanoweki,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tenenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
JOURNAL	Interfering nonneutral evolution from human-chimp-mouse orthologous gene trios				
PUBMED	Science 302 (5652), 1960-1963 (2003)				
REFERENCE	14671302				
AUTHORS	2 (bases 1 to 585)				
TITLE	Clark,A.G., Gnanoweki,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tenenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.				
JOURNAL	Direct Submission				
COMMENT	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA				
FEATURES	These sequences were made by sequencing genomic exons and ordering them based on alignment.				
Source	Location/Qualifiers				
gene	1..585				
	/organism="Mus musculus"				
	/mol_type="Genomic DNA"				
	/db_xref="taxon:10090"				
	<1..>585				
	/gene="AMELX"				
	/locus_tag="HCKM6891"				
ORIGIN					
Query Match	55.0%; Score 41.8; DB 9; Length 585;				
Best Local Similarity	95.6%; Pred. No. 0.00054;				
Matches	43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1 GATTTCCTGCTGCTCTCTGGAGAGACCTTTCCTATGCCCGT 45				
Db	12 GATTTCCTGCTGCTCTCTGGAGAGACCTTTCCTATGCCCGT 56				
RESULT 5					
LOCUS	CD773419	645 bp	mRNA	linear	EST 02-JUL-2003
DEFINITION	AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone				
VERSION	CD773419				
KEYWORDS	IMAGE:30501267 5', mRNA sequence.				
SOURCE	CD773419.1	GI:32431921			
ORGANISM	EST.				
	Mus musculus (house mouse)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 645)				
TITLE	NIH-MGC http://mgs.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Daniela S. Gerhard, Ph.D.				
	Office of Cancer Genomics				
	National Cancer Institute / NIH				
	Bldg. 31 Rm10A07 Bethesda, MD 20892				
	Email: cga@b-rcmail.nih.gov				
	Tissue Procurement: Yoshinobu Yamada, Takashi Nakamura, NIDCR				
	cDNA library Preparation: CLONTECH Laboratories, Inc.				
	cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)				

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
 Plate: NDCM223 row: k column: 04
 High quality sequence stop: 611.
 Location/Qualifiers
 1..645
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30501267"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NH_MGC_190"
 /note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site_1:
 SfiI (ggccatctggcc); Site_2: SfiI (ggcgcccgccg);
 Non-normalized full-length enriched library 5' and 3',
 adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGACCATTTATGACC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGGCGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.71
 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by
 PCR. This library was enriched for full-length clones and
 was constructed by Clontech Laboratories (Palo Alto, CA)
 Corp."

ORIGIN
 Query Match 55.0%; Score 41.8; DB 6; Length 645;
 Best Local Similarity 95.6%; Pred. No. 0.00056;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0,

Oy 1 GATTTCGTTGCTGCTGCTCTGAGGAGCGCTTGTATGCGCCGT 45
 |||||
 Db 84 GATTTCGTTGCTGCTGCTCTGAGGAGCGCTTGTATGCGCCCT 128
 |||||

RESULT 6 707 bp mRNA linear EST 17-JAN-2003
 CB056709 NISC j119G05.w1 Soares NM6P13-15 Mus musculus cDNA clone
 LOCUS IMAGE:4848584 5', mRNA sequence.
 DEFINITION
 CB056709
 CB056709.1 GI:27794996
 EST.
 Mus musculus (house mouse)
 SOURCE
 Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 1 (bases 1 to 707)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Straubeberg, Ph.D.
 Email: cgaps-rcmail.nih.gov
 CDNA Library Preparation:
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LINL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
<http://image.lnl.gov>
 Plate: LHAM10779 row: N column: 9
 Seq primer: T7 primer.
 Location/Qualifiers
 1..707
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:4848584"
 /tissue_type="pituitary gland"
 /dev_stage="juvenile, 13-15 days"
 /lab_host="DH10B (phage-resistant)"

FEATURES
 SOURCE

ORIGIN

/clone.lib="Soares NMBP13-15"
 /note="Organ: Brain; Vector: pT73D-PacI; Site_1: NotI;
 Site_2: EcoRI; 1st strand cDNA was primed with a NotI-
 oligo(dT) primer
 5'-ACTCGAGAGATTCGCGCGCCGCTGATCCGATGTTTTTTTTTTTTTTT-3';
 double-stranded cDNA was ligated to EcoRI adaptors
 5'-ATTTCGACGAGG-3' and 5'-CTCGTCGCG-3' (Pharmacia),
 digested with NotI and cloned into the NotI and EcoRI
 sites of the pT73D-PacI vector. Library went through one
 round of normalization, and was constructed in the
 laboratory of M. Bento Soares (University of Iowa)."

Query Match 55.0%; Score 41.8; DB 6; Length 707;
 Best Local Similarity 95.6%; Pred. No. 0.00057;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GATTTGTTTGCTGCTGCTGCGAGAGAGGCTTTGCTATGCCCGT 45
 |||||||
 58 GATTTGTTTGCTGCTGCTGCGAGAGAGGCTTTGCTATGCCCGT 102

RESULT 7
 AK029358 812 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
 DEFINITION library, clone:4833404E21 product:amelogenin, full insert sequence.
 ACCESSION AK029358.1 GI:26325321
 VERSION AK029358.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE Genome Res. 10 (10), 1617-1630 (2000)
 PUBMED 11042159
 REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, T., Nishi, K., Kitamura, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 11076861
 REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 JOURNAL PANTOM Consortium.
 MEDLINE Functional annotation of a full-length mouse cDNA collection
 PUBMED Nature 409, 685-690 (2001)
 REFERENCE
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 JOURNAL Group Phase I & II Team.
 MEDLINE Analysis of the mouse transcriptome based on functional annotation
 PUBMED of 60,770 full-length cDNAs
 REFERENCE Nature 420, 563-573 (2002)
 PUBMED 6 (bases 1 to 812)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kashiwagi, T.,
 Kato, H., Kawai, D., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Nakazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp,
 URL: http://genome-gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site for further details.
 URL: http://genome-gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers

FEATURES

source

1..812
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:4833404E21"
 /db_xref="taxon:10090"
 /clone="4833404E21"
 /issue_type="head"
 /clone.lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 73..663
 /note="unnamed protein product; amelogenin (MGD) [MG1:88005,
 GI:DJ1768, evidence: BLASTN, 99%, match=799]
 putative"
 /codon_start=1
 /protein_id="BAC26415.1"
 /db_xref="GI:26325321"
 /translation="MGTTILPACILGAARMP,PRHPSGPGYINLSYVLT,PLKXWOS
 MTRQPYSTGYEPMGMLHNOITPVLSQHPSPSHITQPHNHLPVYPAQOPVAPQPM
 PVPGHSMPTQTHQPNIPPSAQOPFQOPFQALPPQSHQPPQSPPLHPQPLAPQ
 PPLPPLFSNQPLSPILPELPLAMPATDKRBEVD"
 793..798
 /note="putative"
 812
 /note="putative"

CDS

polya_signal
 polya_site

ORIGIN

Query Match 55.0%; Score 41.8; DB 3; Length 812;
 Best Local Similarity 95.6%; Pred. No. 0.00057;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GATTTGTTTGCTGCTGCTGCGAGAGAGGCTTTGCTATGCCCGT 45
 |||||||
 84 GATTTGTTTGCTGCTGCTGCGAGAGAGGCTTTGCTATGCCCGT 128

RESULT 8
 CBS588212 843 bp mRNA linear EST 03-APR-2003
 LOCUS AGENCOURT 12771359 NIH MGC 136 Mus musculus cDNA clone
 DEFINITION IMAGE:30293239 5', mRNA sequence.
 ACCESSION CBS588212
 VERSION CBS588212.1 GI:29506068
 KEYWORDS EST.

SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE AUTHORS TITLE JOURNAL COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (baae 1 to 843) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: http://image.llnl.gov Plate: NDAM338 row: 9 column: 08 High quality sequence strip: 636. Location/Qualifiers 1. 843 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:30293239" /tissue_type="embryonic limb, maxilla and mandible" /lab_host="DH10B (phage-resistant)" /clone_1lb="NIH MGC 136" /note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible; embryonic day 17.5, 18.5 and newborn (mandible (5), 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTGTTCTGATCGCAGCGCCGCCCT(7)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue constructed by David Rowe; library constructed by Reagen, Invitrogen Corp. Note: this is a NIH_MGC Library."
ORIGIN	Query Match 55.0%; Score 41.8; DB 6; Length 843; Best Local Similarity 95.6%; Pred. No. 0.0006; Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0.
Dn	1 GATTTTGTTGGCTCCTCTGGAGAGACGCTTTCCTATGCCCGT 45 51 GATTTTGTTGGCTCCTCTGGAGAGACGCTTTCCTATGCCCGT 95
RESULT 9	CB588525 859 bp mRNA linear EST 03-APR-2003
LOCUS	AGENCOURT 12567944 NIH MGC 136 Mus musculus cDNA clone
DEFINITION	IMAGE:30289751 5', mRNA sequence.
ACCESSION	CB588525
VERSION	CB588525.1 GI:29506381
KEYWORDS	EST.
SOURCE ORGANISM	Mus musculus (house mouse)
REFERENCE AUTHORS TITLE JOURNAL COMMENT	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (baae 1 to 859) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at:

```

http://image.lnl.gov
plate: NDAM329 row: e column: 24
High quality sequence stop: 763.
Location/Qualifiers
1. .859
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="NIH MGC 136"
/notes="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACAGTTCGTAAGCGAGCGCGCCCT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 859;
Best Local Similarity 95.6%; Pred. No. 0.0006;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GATTTTGTCCTGCCTCCCTGGAGAGAGCCTTGTCATGCCCGT 45
|||||
70 GATTTTGTCCTGCCTCCCTGGAGAGAGCCTTGTCATGCCCGCT 114

RESULT 10
LOCUS CB589177
DEFINITION AGENCOURT_12770446 NIH_MGC_136 Mus musculus cDNA clone
ACCESSION CB589177
VERSION CB589177.1 GI:29507033
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 865)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@db-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDAM31 row: f column: 15
High quality sequence stop: 630.
Location/Qualifiers
1. .865
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290534"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="NIH MGC 136"
/notes="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw

```

equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 865;
Best Local Similarity 95.6%; Pred. No. 0.0006;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 45
|||||
Db 69 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 113
|||||

RESULT 11
LOCUS CB587051
DEFINITION AGENCOURT 12807074 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30295123 5', mRNA sequence.
ACCESSION CB587051
VERSION CB587051.1 GI:29504907
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 886)
NH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM343 row: e column: 20
High quality sequence stop: 619.
Location/Qualifiers
1. .886
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30295123"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

FEATURES
source

1. .886
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30295123"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 865;
Best Local Similarity 95.6%; Pred. No. 0.0006;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 45
|||||
Db 95 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 139
|||||

RESULT 12
LOCUS CB589251
DEFINITION AGENCOURT 12780700 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30287069 5', mRNA sequence.

ACCESSION CB589251
VERSION CB589251.1 GI:29507107
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 889)
NH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM322 row: f column: 06
High quality sequence start: 23
High quality sequence stop: 581.
Location/Qualifiers
1. .889
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30287069"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

FEATURES
source

1. .889
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30287069"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 55.0%; Score 41.8; DB 6; Length 889;
Best Local Similarity 95.6%; Pred. No. 0.0006;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 45
|||||
Db 89 GATTGTGTTGCTGCTCTCTGAGAGAGCCTTTGCTATGCCCGT 133
|||||

RESULT 13
LOCUS CB574837
DEFINITION AGENCOURT 12973279 NIH_MGC_136 Mus musculus cDNA clone
IMAGE:30254719 5', mRNA sequence.

ACCESSION CB574837
VERSION CB574837.1 GI:29494367
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 891)
NH-MGC <http://mgc.nci.nih.gov/>.

REFERENCE
AUTHORS

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: NDM342 row: d column: 24
 High quality sequence start: 19
 High quality sequence stop: 658.
FEATURES Location/Qualifiers
 1..891
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30294719"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /clone_1b="NIH_MGC_136"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 55.0%; Score 41.8; DB 6; Length 891;
 Best Local Similarity 95.6%; Pred. No. 0.0006;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCCTGCTCTGGAGAGAGCGCTTGTATGCCCGT 45
 |||||
 90 GATTTGTTTGCCTGCTCTGGAGAGAGCGCTTGTATGCCCGT 134

RESULT 14
 CBS90451 913 bp mRNA linear EST 03-APR-2003
 LOCUS AGENCOURT_12770290 NIH_MGC_136 Mus musculus CDNA clone
 IMAGE:30286944 5', mRNA sequence.
 CBS90451
 CBS90451.1 GI:29508307
 EST.
 Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 913)
 NIH-MGC <http://imgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: NDM322 row: a column: 01
 High quality sequence start: 12
 High quality sequence stop: 640.
FEATURES Location/Qualifiers

source
 1..913
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30286944"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /clone_1b="NIH_MGC_136"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 55.0%; Score 41.8; DB 6; Length 913;
 Best Local Similarity 95.6%; Pred. No. 0.00061;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCCTGCTCTGGAGAGAGCGCTTGTATGCCCGT 45
 |||||
 82 GATTTGTTTGCCTGCTCTGGAGAGAGCGCTTGTATGCCCGT 126

RESULT 15
 CBS87332 919 bp mRNA linear EST 03-APR-2003
 LOCUS AGENCOURT_12971606 NIH_MGC_136 Mus musculus CDNA clone
 IMAGE:30291971 5', mRNA sequence.
 CBS87332
 CBS87332.1 GI:29505188
 EST.
 Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 919)
 NIH-MGC <http://imgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Plate: NDM335 row: b column: 12
 High quality sequence start: 216
 High quality sequence stop: 325.
FEATURES Location/Qualifiers
 1..919
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30291971"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /clone_1b="NIH_MGC_136"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGAGCGGCCGCTT)15-3'. Size selected for the >1kb fragments, average insert size

ORIGIN

1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe, library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

Query Match	55.0%	Score 41.8	DB 6	Length 919
Best Local Similarity	95.6%	Pred. No. 0.00061		
Matches	43	Conservative	0	Mismatches 2; Indels 0; Gaps 0;
Qy	1	GATTTCCTTTCCTCCCTCCGAGAGACCTTCGATGCGCGT	45	
Db	69	GATTTCCTTTCCTCCCTCCGAGAGACCTTCGATGCGCGT	113	

Search completed: February 8, 2005, 22:18:06
Job time : 906.232 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 390.297 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70
Sequence: 1 gattctgttcgtcgtccccc.....aaattcccaattcccaattc 70

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.8	71.1	6451	AB091789	AB091789 Bos tauri
2	49.2	70.3	363	11	BV0882934 RPAMSE00
3	49.2	70.3	363	11	BV097602 RPAMSE00
4	49.2	70.3	463	10	DB3063S2 Mus musculi
5	49.2	70.3	695	11	BV089295 RPAMSE00
6	49.2	70.3	695	11	BV097603 RPAMSE00
7	49.2	70.3	9384	10	AF294397 Mus muscu
8	49.2	70.3	95826	10	AL805974 Mouse DNA
9	47.6	68.0	212886	2	AC093946 Rattus no
10	47.6	68.0	234471	2	AC121424 Rattus no
11	46.8	66.9	8004	9	AB091786 Lemur cat
12	46.8	65.7	7425	4	AB091791 Sus scrofa
13	45	64.3	6442	9	AB091783 Salimiri b
14	43.6	62.3	5562	9	AB091787 Oculimur
15	43.6	62.3	5564	9	AB091785 Lemur cat
16	43.6	62.3	6465	9	AB091781 Pan trogl
17	43.6	62.3	8810	9	AY040206 Homo sapi
18	43.6	62.3	158142	2	AL357130 Homo sapi
19	43.6	62.3	259202	9	AC002366 Human Xp2

20	42.8	61.1	7163	9	AB091782 Pan trogl
21	42.8	61.1	38765	2	BS000568 Pan trogl
22	42.8	61.1	177654	2	AP000918 Homo sapi
23	42.8	61.1	177726	9	AC145770 Pan trogl
24	42.8	61.1	190089	9	BS000576 Pan trogl
25	42.8	61.1	200214	9	AC013412 Homo sapi
26	41.8	59.7	6264	4	AB091790 Bos tauri
27	40.8	58.3	5712	4	AB091793 Equus cab
28	40.2	57.4	746	4	AF215880 Capra hie
29	40.2	57.4	752	4	AF215889 Capra hie
30	40.2	57.4	757	4	BOVAMLGNI
31	40.2	57.4	820	4	BOVAMLGNI
32	38.6	55.1	778	4	SSU43405
33	38.6	55.1	6831	4	AB091792 Sus scrofa
34	37	52.9	7454	4	AB091784 Salimiri b
35	35.8	51.1	5591	4	AB091794 Equus cab
36	35.6	50.9	296	10	MSNAMEB
37	35.6	50.9	727	10	MSNAMEB
38	35.6	50.9	765	4	AB032194 Equus cab
39	35.6	50.9	789	4	AB032193 Equus cab
40	35.6	50.9	798	10	BC053090 Mus muscu
41	35.6	50.9	799	10	MSNAMEA
42	35.6	50.9	825	6	AX865705 Sequence
43	35.6	50.9	825	6	BD145767 Primer fo
44	35.6	50.9	1626	6	AX882160 Sequence
45	35.6	50.9	1626	6	BD159567 Primer fo

ALIGNMENTS

RESULT 1	AB091789	AB091789	AMELX gene for amelogenin, partial cds.
LOCUS	AB091789	6451 bp	DNA linear MAM 02-MAY-2003
DEFINITION	Bos taurus		
ACCESSION	AB091789		
VERSION	AB091789.1	GI:29126030	
KEYWORDS			
SOURCE	Bos taurus (cow)		
ORGANISM	Bos taurus		
REFERENCE	Iwase,M., Satra,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.		
AUTHORS	Iwase,M., Satra,Y. and Takahata,N.		
TITLE	Direct Submision		
JOURNAL	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryuwo1.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)		
MEDLINE	22608569		
PUBMED	12672962		
REFERENCE	2 (bases 1 to 6451)		
AUTHORS	Iwase,M., Satra,Y. and Takahata,N.		
TITLE	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shonan kokusaiimura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryuwo1.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)		
FEATURES	Location/Qualifiers		
source	1..6451		
gene	/organism="Bos taurus"		
5'UTR	/mol_type="genomic DNA"		
CDS	/db_xref="taxon:9913"		
	/sex="male"		
	238..6451		
	/gene="AMELX"		
	join(238..340,1650..1661)		
	/gene="AMELX"		
	join(1662..1715,4173..4220,5560..5601,5665..5709,5987..6451)		
	/gene="AMELX"		
	/codon_start=1		
	/product="amelogenin"		

JOURNAL DNA Seq. 6 (5), 307-310 (1996)
 MEDLINE 97142134
 PUBMED 8988368
 REFERENCE 2 (bases 1 to 463)
 AUTHORS Oida, S.
 JOURNAL Direct Submission
 Submitted (14-JAN-1996) Shinichiro Oida, Tokyo Medical and Dental University, Fac. Dentistry, Dept. Biochemistry, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (E-mail: s.oida.opchdent.tmd.ac.jp, Tel: 03-5803-5448, Fax: 03-5803-0187)
 On Dec 15, 1997 this sequence version replaced gi:2662357.
 FEATURES
 source
 1..463
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /sub_species="Balb-c"
 /db_xref="taxon:10090"
 /chromosome="X"
 /map="XF3-F4"
 /clone="MAMEG-2"
 /cell_type="Leukocyte"
 /clone_lib="EMBL-3 SP6-T7"
 /dev_stage="adult"
 /join(D83063.1:203..258,222..233)
 <1..221
 /number=1
 222..287
 /product="amelogenin"
 /number=2
 288..463
 /number=2
 /number=2
 ORIGIN
 Intron
 Query Match 70.3%; Score 49.2; DB 10; Length 463;
 Best Local Similarity 81.4%; Pred. No. 1.7e-08;
 Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGTAGTAAATTTCCAA 60
 DB 245 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGTAGTAAATTTCCAA 304
 QY 61 TTTCATTT 70
 DB 305 TTACTTATT 314
 Db
 RESULT 5
 BV089295/c 695 bp DNA linear STS 15-OCT-2003
 LOCUS RPAWMSB00001191 Roche Palo Alto Mus musculus STS genomic, sequence
 DEFINITION tagged site.
 ACCESSION BV089295
 VERSION BV089295.1 GI:37666774
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 695)
 Uenka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.
 Mus musculus SNPs
 Unpublished (2003)
 TITLE
 JOURNAL
 COMMENT
 Contact: Jonathan Uenka
 Roche Palo Alto Genetics and Genomics Department
 Roche Palo Alto
 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
 Tel: 6508555807
 Email: Jonathan.Uenka@roche.com
 Primer A: No primer submitted
 Primer B: No primer submitted.
 Location/Qualifiers
 FEATURES

source
 1..695
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="X-73613-72919-AL805974.7.1.95826"
 /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/Heu, 129/Sv, AKR/J, B10.D2-H2/oshu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MPJ, NZB/B1mJ, NZW/Lac, SPRET/Ei."
 <1..>695
 ORIGIN
 STS
 Query Match 70.3%; Score 49.2; DB 11; Length 695;
 Best Local Similarity 81.4%; Pred. No. 1.8e-08;
 Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGTAGTAAATTTCCAA 60
 DB 479 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGTAGTAAATTTCCAA 420
 QY 61 TTTCATTT 70
 DB 419 TTACTTATT 410
 Db
 RESULT 6
 BV097603 695 bp DNA linear STS 14-JAN-2004
 LOCUS RPAWMSB00010113 Roche Palo Alto Mus musculus STS genomic, sequence
 DEFINITION tagged site.
 ACCESSION BV097603
 VERSION BV097603.1 GI:40810679
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 695)
 Uenka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J.D., Foerzler, D. and Peltz, G.
 Mus musculus SNPs
 Unpublished (2003)
 TITLE
 JOURNAL
 COMMENT
 Contact: Jonathan Uenka
 Roche Palo Alto Genetics and Genomics Department
 Roche Palo Alto
 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
 Tel: 6508555807
 Email: Jonathan.Uenka@roche.com
 Primer A: No primer submitted with this STS
 Primer B: No primer submitted with this STS.
 Location/Qualifiers
 FEATURES
 source
 1..695
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="X-73613-72919-AL805974.7.1.95826"
 /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/Heu, 129/Sv, AKR/J, B10.D2-H2/oshu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRL/MPJ, NZB/B1mJ, NZW/Lac, SPRET/Ei."
 <1..>695
 ORIGIN
 STS
 Query Match 70.3%; Score 49.2; DB 11; Length 695;
 Best Local Similarity 81.4%; Pred. No. 1.8e-08;
 Matches 57; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 1 GATTTGTTGCTGCTCTGAGACAGCCTTCAGTATCCCTGTAGTAAATTTCCAA 60

Db 479 GATTTTGTTCGCTGCTCCTGGAGACACTTTTGCATGCCCGAATTAATCAATCCA 420

Qy 61 TTTCGAATTT 70
||| |
Db 419 TTCTACTAATT 410

RESULT 7	AF294397	9384 bp	DNA	linear	ROD 03-SEP-2000
LOCUS	AF294397				
DEFINITION	Mus musculus amelogenin gene, promoter and partial cds.				
ACCESSION	AF294397				
VERSION	AF294397.1	GI:3965405			
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Euhetia; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 3384)
AUTHORS	Sneed,M.L., Paine,W.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P., Paine,C.T., Chen,L.S., Burstein,J.M., jitpdeeinditra,S., White,S.N. and Bringas,P. Jr.
TITLE	Transgene animal model for protein expression and accumulation into forming enamel
JOURNAL	Connect. Tissue Res. 38 (1-4), 279-286 (1998)
PMID	9615044

REFERENCE	2 (bases 1 to 9384)
AUTHORS	Sneed, M.L., Zhu, D.-H., Lei, Y.-P. and Paine, M.L.
TITLE	Direct Submission
JOURNAL	Submitted (08-AUG-2000) Dentistry, University of Southern California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES	
SOURCE	1 location/Qualifiers
	8284

```

source
    1..9384
        /locus_tag="Mmu116L07"
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="X"
promoter
    1..8546
TATA_signal
    7168..7194
mRNA
    8547..8612
        /product="amelogenin"
CDS
    8559..8612
```

ORIGIN
/translation="MGTWILFACLLGAAPAMP"

Query Match	70.3%	Score 49.2	DB 10	Length 9584
Best Local Similarity	81.4%	Pred. No. 2	7E-08	
Matches	57	Conservative	0	Mismatches 13; Indels 0; Gaps 0;

QY	1	GATTTTGTTCCTCTGCGAGAGCCTCAGATACCTCTGTGATTAATTTCCAA	60
Db	8570	GATTTTGTTCCTCTGCGAGAGCCTTTGCTATCCCGGTAGTAAACAATCCA	8629
QY	61	TTTCCAAATTT	70
Db	8630	TTTACTAATTT	8639

RESULTS	AL805974	95826 bp	DNA	linear	ROD 29-AUG-2003
LOCUS	Mouse DNA sequence from clone RP23-334F21				on chromosome X, complete
DEFINITION	sequence,				
ACCESSION	AL805974				
VERSION	AL805974.8				
KEYWORDS	HTG.	GI:34366495			

SOURCE ORGANISM	REFERENCE AUTHORS TITLE JOURNAL
Mus musculus (house mouse)	Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK
Mus musculus	Chapman, J. Direct Submission
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus	Submitted (19-SEP-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK
1 (bases 1 to 95826)	Chapman, J. Direct Submission

COMMENT On Aug 30, 2003 this sequence version replaced gi:25955748. Sequences from the Vector Ontology Consortium.

TITLE Direct Submission
JOURNAL Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT Clone requests: clonerequest@sanger.ac.uk
 On Aug 30, 2003 this sequence version replaced g1:25955748.
 Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternated chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em1, EMBL; Sw1, SWISSPROT; Tr1, TREMBL; Wp1, WORMPEP; Information from the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-334F21 is from the RPCI-23 Mouse BAC library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACes1.6.

```

FEATURES
  source
    Location/Qualifiers
      1..95826
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /chromosome="X"
        /clone="RP23-334P21"
        /clone_lib="RPCT-23"

```

[illegible]

RESULT 9					
AC093946/c					
LOCUS	AC093946	212886 bp	DNA	linear	HTG 10-MAY-200
DEFINITION	Rattus norvegicus clone CH230-166F2, *** SEQUENCING IN PROGRESS				
ACCESSION	AC093946				

VERSION
 AC093946.9 GI:30520433
 HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS
 Rattus norvegicus (Norway rat)
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 212886)
 REFERENCE
 Muzny, D., Maric, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Albrooks, S., Amin, A., Arguliano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, B.,
 Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Fleggs, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hu, Y., Hume, J., Idelbird, D., Jackson, A.,
 Jackson, J., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowalski, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, A.,
 Lorenzen, L., Louie, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Mlosovskiy, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokedi, O., Okunolu, G., Olarnpunsoon, A., Pal, S., Parks, K.,
 Pastermak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polinder, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivas, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaiz, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Soze, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanil, K.,
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wood, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhou, D., von
 Niederhuesern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 TITLE
 Unpublished
 JOURNAL
 2 (bases 1 to 212886)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (12-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 212886)
 REFERENCE
 Rat Genome Sequencing Consortium.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT
 On May 10, 2003 this sequence version replaced gi:2365261.
 The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GPVN
 Center clone name: CH230-166F2
 ----- Summary Statistics
 Assembly program: Atlas 3.0i
 Consensus quality: 207370 bases at least Q40
 Consensus quality: 208312 bases at least Q30
 Consensus quality: 208768 bases at least Q20
 Estimated insert size: 218822; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 212886: contig of 212886 bp in length.
 Location/Qualifiers
 1. 212886
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-166F2"
 1. 1230
 /notes="wgs contig"
 211449..212886
 /note="wgs_contig"
 ORIGIN
 Query Match 68.0%; Score 47.6; DB 2; Length 212886;
 Best Local Similarity 80.0%; Pred. No. 1.8e-07;
 Matches 56; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 1 GATTTGTTTGCTGCTGCTGAGACAGCTTCACTATGCTGTGAGTAATTTCCAA 60
 DB 112447 GATCTGTGTGCTGCTGCTGAGACAGCTTGTCTATGCGCCGTGATTAACACCCCA 112388
 QY 61 TTTCCAAATT 70
 DB 112387 TTTACTGATT 112378
 RESULT 10
 AC121424 234471 bp DNA linear HTG 21-SEP-2002
 AC121424
 DEFINITION
 Rattus norvegicus clone CH230-122F17, *** SEQUENCING IN PROGRESS
 AC121424
 AC121424
 ACCESSION
 AC121424.3 GI:2365350
 VERSION
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS
 Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

- 1 (bases 1 to 234471)
Munzy,D,Marle, Metzker,M,lee, Abramson,S, Adams,C, Alder,D, Allen,C, Allen,H, Albrooks,S, Amin,A, Angiano,D, Anyalbech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Blawie,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabies,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haaland,W, Hamli,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howell,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,O, Lorenshewa,L, Louieged,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E, McWhiney,S, McLeod,M,P, McNeill,T,Z, Meenan,E, Mlosoavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwaokwelen,O, Okunolu,G, Olarnpungsoon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plapper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L, L, Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reich,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,U, Shvartbeyn,A, Sisson,I, Sitter,C,D, Smays,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabors,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Umani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhou,S, Dunn,D, Von Niederhausen,A, Weiser,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstein,G, and Gibbs,R,A.
- Direct Submission
Unpublished
2 (bases 1 to 234471)
Worley,K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 234471)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Sep 21, 2002 this sequence version replaced gi:21908624. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be

contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXUI

Center clone name: CH230-122F17

Summary Statistics

Assembly program: Phrap, version 0.990329

Consensus quality: 220465 bases at least Q40

Consensus quality: 222988 bases at least Q30

Consensus quality: 224416 bases at least Q20

Estimated insert size: 241282; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a "working draft" sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 232989: contig of 232989 bp in length

* 232990 233089: gap of unknown length

* 233090 234471: contig of 1382 bp in length.

FEATURES

source

1. 234471

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-122F17"

104912..105798

/note="clone boundary"

clone_end:77

site:ECORI

end_sequence:BH268493"

complement(231872..232742)

/note="clone boundary"

clone_end:Sp6

site:ECORI

end_sequence:BH268494"

ORIGIN

Query Match 68.0%; Score 47.6; DB 2; Length 234471;

Best Local Similarity 80.0%; Pred. No. 1.8e-07;

Matches 56; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY

1 GATTTGTTGTCCTGCTCTGGAGCAGCCTTCAGTATGCTGAGTAATTTCCAA 60

DB 161729 GATCTGTTGTCCTGCTCTGGAGCAGCCTTTGCTATGATCGCTGATTAACCCCA 161788

QY

61 TTTCGAATT 70

DB 161789 TTACTGATT 161798

RESULT 11

AB091786

LOCUS AB091786 8004 bp DNA linear PRI 02-MAY-2003

DEFINITION Lemur catta AMELY gene for amelogenin, partial cds.

AB091786

ACCESSION AB091786.1 GI:29126025

KEYWORDS

SOURCE

ORGANISM Lemur catta (ring-tailed lemur)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

	gene	/organism="Sus scrofa"
		/mol_type="genomic DNA"
		/db_xref="taxon:9823"
		/sex="male"
		1491..7425
		/gene="AMELX"
		join(1491..1548,3140..3351)
		/gene="AMELX"
		join(3152..3205,5316..5363,6619..6660,6717..6761,
		7033..7425)
		/gene="AMELX"
		/codon_start=1
		/product="amelogenin"
		/protein_id="BAC6110.1"
		/db_xref="GI:29126035"
		/translation="MGTWILPACILCALGAFSPMPPIPHGHQYINFSVEDIYLEAIRID
		RFAVLPLPKMYOMNLRPYTSGYGEPMGWMLHQQIIPVVSQOTPPSHALQPHNIIM
		VPAQGPDPQPMWPLRGHSMTPTQHQNLPPLPAQOPPOPOPVPORPQRPLOPSBP
		MHPIQPLPQPPLPPMPSMQSLPLDPLEAMPAT"
ORIGIN		
Query Match	65.7%; Score 46; DB 4; Length 7425;	
Beech Local Similarity	83.9%; Pref. No. 5e-07;	
Matches	52; Conservative 0; Mismatches 10; Indels 0; Gaps 0;	
Oy	1 GATTTCCTTGCCCGCCTCCGGAGCAGCCCTTCAGTGCCTGTGATAAATTTCCA 60	
Dd	3163 GATTTCCTTGCTGCCCTCCGGAGCAGCCCTTCATGCTCTTAAGTAAAGCACCCC 3222	
Oy	61 TT 62	
Dd	3223 TT 3224	
RESULT_13		
LOCUS	ABO91783 6442 bp DNA linear PRI 02-MAY-2003	
DEFINITION	Saimiri sciureus AMELX gene for amelogenin, partial cds.	
ACCESSION	ABO91783	
VERSION	ABO91783.1 GI:29126019	
KEYWORDS		
SOURCE		
ORGANISM	Saimiri sciureus (common squirrel monkey)	
	Saimiri sciureus	
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;	
	Saimiri.	
REFERENCE	1	
AUTHORS	Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Inai,H. and Takahata,N.	
TITLE	From the Cover: The amelogenin loci span an ancient pseudautosomal	
JOURNAL	boundary in diverse mammalian species	
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)	
PUBMED	23608569	
REFERENCE	12672962	
AUTHORS	2 (bases 1 to 6442)	
TITLE	Iwase,M., Satta,Y. and Takahata,N.	
JOURNAL	Direct Submission	
	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for	
	Advanced Studies(Sokenдай), Department of Biosystems Science;	
	Shonan Kokusaimitura, Hayama, Kanagawa 240-0193, Japan	
	(E-mail:iwasemin@koryuw01.soken.ac.jp, Tel:81-468-58-1571,	
	Fax:81-468-58-1544)	
FEATURES	Location/Qualifiers	
Source	1..6442	
	/organism="Saimiri sciureus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9521"	
	/sex="male"	
	936..6442	
	/gene="AMELX"	
	join(936..991,2298..2309)	
	/gene="AMELX"	
	join(2310..2363,4289..4336,5592..5633,5725..5769,	
	6041..>6442)	

```

/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="HACE103.1"
/db_xref="GI:29126020"
/translation="MGTWILFLPCLLGAFAFAMPLRPHRPGPGYINFSYENSHSQAINIDIRGLVLRPLKYGOSMIRIPAYSPGLYRPMGWNHQQIILPVLSQDRPTTLQPHNHIPVLRVPAQDPVPOQPMHVPQGSQSMIPRONHQPMLPRPAQDPFQFQVQDPQHQMOPQEPFVHNPQRLRQRRRLPRMFMQRLPMLPDLRLAAMDAT"

```

Query Match	64.3%	Score 45;	DB 9;	Length 6442;
Best Local Similarity	90.6%	Pred. No. 1.3e-06;		
Matches 48;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;

Oy		1	GATTTGTTTTCTGCCTCCTGGAGACAGCTTCAGTATGCCGTGAGTAATAA	53
Db		2321	GATTTTGTTTCCCTGCCTCCTGGAGACAGCTTTTGGCATGCCTGTGACTAAAA	2373

RESULT 14	AB091787	5562 bp	DNA	linear	PRI 02-MAY-2003
LOCUS	AB091787				
DEFINITION	Otolemur garnettii AMELX gene for amelogenin, partial cds.				
ACCESSION	AB091787				
VERSION	AB091787.1	GI:29126027			
KEYWORDS					
SOURCE	Otolemur garnettii (small-eared galago)				
ORGANISM	Otolemur garnettii				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1
Iwase,M., Saita,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
2608559
2 (bases 1 to 5562)
12672962
Iwase,M., Saita,Y. and Takahata,N.
Direct Submission Mineyo Iwase, Graduate University for
Submitted (19-Sep-2002) International Science

FEATURES

```

gene      267. .5562
          /gene="AMELX"
5'UTR     join(267. .321,1401. .1412)
          /gene="AMELX"
CDS       join(1413. .1466,3385. .3432,4599. .4740,4832. .4876

```

ORIGIN

Query Match	62.3%	Score 43.6	DB 9	Length 5562
Similarity	84.5%	Pred. No. 4.5e-06		
Matches 49; Conservative	0	Mismatches 9	Indels 0	Gaps 0

Dy 1 GATTTGTGTTGCCCTGCCTCCGGAGACAGCCTTCAGTAGTGCGTAGTAATTTCC 58

Db 1424 GATTTGTGTTGCCCTGCCTCCGGAGACAGCCTTTGCCATGCCCCGTGAGTAAACACC 1461

RESULT 15	AB091785	AB091785	5684 bp	DNA	linear	PRI 02-MAY-2003
LOCUS	AB091785					
DEFINITION	Lemur catta AMELX gene for amelogenin, partial cds.					
ACCESSION	AB091785					
VERSION	AB091785.1	GI:29126023				
KEYWORDS						
SOURCE	Lemur catta (ring-tailed lemur)					
ORGANISM	Lemur catta					

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudoautosoma-
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
22608569
12672962
2 (bases 1 to 5684)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (13-Sep-2002) Mineyo Iwase, Graduate University for
Advanced Science and Technology

```

source
1. .5684
/organism="Lemur catta"
/mol_type="genomic DNA"
/db_xref="taxon:9447"
/sex="male"
100. .5684
gene
100. .5684

```

```

CDS                                  /gene="AMELX"
5229..5568) join(1515..1568,3465..3512,4804..4845,4937..4981

```

ORIGIN

Query Match	62.3%	Score 43.6	DB 9	Length 5684
Best Local Similarity	84.5%	Pred. No. 4.5e-06		
Matches 43	Conservative	0	Mismatches 9	Indels 0
				Gaps 0

Oy 1 GATTTTGTGGCTGCTCTGGGAGCAGCCTTAGTATGCTGTGAGTAAATTTCC 58
 |||||
 Db 1526 GATTTTGTGGCTGCTCTGGGAGCAGCCTTAGTATGCTGTGAGTAAACACCC 1583

Search completed: February 8, 2005, 16:36:55
Job time : 391.297 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 : Search time 92.858 seconds
(without alignment) 3957.220 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70
Sequence: 1 gattctgttcgctgcctcc.....aaattccattcccaatt 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseq1980s:*
2: geneseq1990s:*
3: geneseq2000s:*
4: geneseq2001s:*
5: geneseq2001bs:*
6: geneseq2002as:*
7: geneseq2002bs:*
8: geneseq2003as:*
9: geneseq2003bs:*
10: geneseq2003cs:*
11: geneseq2003ds:*
12: geneseq2004s:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.6	62.3	259202	12	ADQ18492 Human sof
2	40.2	57.4	722	6	AA141111
3	40.2	57.4	752	6	AA141110
4	38	54.3	270	12	ADM80835
5	38	54.3	549	12	ADM80826
6	38	54.3	623	12	ADM80827
7	35.6	50.9	825	4	AAH03775
8	35.6	50.9	1626	4	AAH17575
9	35.6	50.9	2176	4	AAH17101
10	35.6	50.9	3037	12	ADQ24117
11	34.6	49.4	556	10	ADM59026
12	34.6	49.4	556	10	ADM53782
13	34	48.6	65	6	ABN28369
14	34	48.6	476	6	AAZ50832
15	33.8	48.3	318	12	ADM80836
16	33.8	48.3	750	2	AAZ07020
17	33.8	48.3	793	2	AAZ07018
18	32.2	46.0	802	2	AAZ07019
19	32.2	46.0	852	12	ADQ22977
20	28.6	40.9	277	3	AAC55907
21	26.4	37.7	389	5	AAFe4393

C	22	26.4	37.7	423	4	ABA58214	Abas58214 Human foe
C	23	26.4	37.7	423	4	AAI37826	AAI37826 Probe #65
C	24	26.4	37.7	423	4	AAK31962	AAK31962 Human bon
C	25	26.4	37.7	423	4	AAK06298	AAK06298 Human bra
C	26	26.4	37.7	423	4	AB831653	AB831653 Human liv
C	27	26.4	37.7	423	6	AB806725	AB806725 Human gen
C	28	26.4	37.7	567	4	AA143405	AA143405 Probe #12
C	29	26.4	37.7	941	2	AAV84413	AAV84413 Human sec
C	30	26.4	37.7	941	2	AAV84618	AAV84618 Human sec
C	31	26.4	37.7	941	4	ABA83196	ABA83196 Human sec
C	32	26.4	37.7	941	4	ABA83401	ABA83401 Human sec
C	33	26.4	37.7	941	9	ACH04902	ACH04902 Novel hum
C	34	26.4	37.7	941	9	ACH04697	ACH04697 Novel hum
C	35	26.4	37.7	941	9	ACD44712	ACD44712 Human CDN
C	36	26.4	37.7	941	9	ACD44507	ACD44507 Human CDN
C	37	25.8	36.9	292	3	ACC28078	ACC28078 Human sec
C	38	25.8	36.9	2000	10	ACC61018	ACC61018 Gene sequ
C	39	25.8	36.9	2000	10	ADK62663	ADK62663 Disease t
C	40	25.4	36.3	199878	10	ADL13719	ADL13719 Osteoarth
C	41	25.2	36.0	1639	12	ADN05617	ADN05617 Antipsoit
C	42	25.2	36.0	48436	6	ABN89533	ABN89533 Human cor
C	43	25.2	36.0	96588	9	ADA03026	ADA03026 Human MEN
C	44	25.2	36.0	96588	10	ADB72764	ADB72764 Human MEN
C	45	25.2	36.0	96588	10	ADC85506	ADC85506 Human MEN

ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
ID	ADQ18492	
XX	AC	ADQ18492;
XX	AC	
XX	DT	26-AUG-2004 (first entry)
XX	DB	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.
XX	KW	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; ds.
XX	KW	
XX	OS	Homo sapiens.
XX	XX	
XX	PD	10-JUN-2004.
XX	PF	26-NOV-2003; 2003WO-US038193.
XX	PR	26-NOV-2002; 2002US-0429739P.
XX	XX	
XX	PA	(PROT-) PROTEIN DESIGN LABS INC.
XX	PI	Aziz N, Ginsburg WM, Zlotnik A;
XX	DR	WPI; 2004-441208/41.
XX	XX	
XX	PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample
XX	PT	and comparing the gene expression, also useful in treating soft tissue sarcoma.
XX	XX	
XX	PS	Example 2; SEQ ID NO 1311; 210bp; English.
XX	XX	
XX	CC	The invention relates to a novel method for detecting soft tissue sarcoma
XX	CC	which comprises obtaining a first soft tissue sample from an individual
XX	CC	and a normal soft tissue sample from the same or different individual,
XX	CC	determining the expression of a gene in both samples and comparing the
XX	CC	expression of the gene in both soft tissue samples, where a higher level
XX	CC	of protein expression in the first soft tissue sample indicates the
XX	CC	presence of soft tissue sarcoma. The method of the invention has
XX	CC	cytostatic applications and may be useful for detecting soft tissue

CC cancer, obesity and Tangier disease.
XX
SQ Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Query Match 54.3%; Score 38; DB 12; Length 549;
Best Local Similarity 75.8%; Pred. No. 0.00045;
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTGCTGAGACAGCTTCAGTATGCTGTGAGTAAATTTCCA 60
DB 47 GATTTATTTGCTGCTGCTGAGACAGCTTTGCCATGCTGTGCTTACCCCTTTGAA 106
QY 61 TT 62
DB 107 GT 108
RESULT 6
ADM80827
ID ADM80827 standard; cDNA; 623 BP.
AC ADM80827;
XX
XX 03-JUN-2004 (first entry)
DT
DE Human CADEC14 encoding cDNA SEQ ID NO:56.
XX
XX human; cell adhesion and extracellular matrix protein; CADEC14;
XX neuroprotective; cytoskeletal; anorectic; immune disorder;
XX connective tissue disorder; developmental disorder;
XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX Tangier disease; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 36..563
FT /*tag= a
FT /product= "CADEC14"
FT
XX
XX W02004015396-A2.
XX
XX 19-FEB-2004.
XX
XX 12-AUG-2003; 2003WO-US025418.
XX
XX 13-AUG-2002; 2002US-0403781P.
XX 30-AUG-2002; 2002US-0407034P.
XX 13-SEP-2002; 2002US-0410566P.
XX 24-SEP-2002; 2002US-0413482P.
XX 25-SEP-2002; 2002US-0413890P.
XX 08-NOV-2002; 2002US-0424504P.
XX 13-NOV-2002; 2002US-0426222P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Ellicott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J, Blake JJ;
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Wang UT;
XX Wang UT, Chien D, Yang YG;
XX
XX WPI; 2004-191795/18.
XX P-PSDB; ADM80785.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.
XX
XX Claim 5; SEQ ID NO 56; 272pp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
XX matrix protein designated CADEC14. CADEC14 sequences has neuroprotective,
XX cytoskeletal and anorectic activities. The CADEC14 polypeptides and

CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX neurological, developmental, connective tissue and cell proliferative
XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX cancer, obesity and Tangier disease.
SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
Query Match 54.3%; Score 38; DB 12; Length 623;
Best Local Similarity 75.8%; Pred. No. 0.00047;
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTGCTGAGACAGCTTCAGTATGCTGTGAGTAAATTTCCA 60
DB 47 GATTTATTTGCTGCTGCTGAGACAGCTTTGCCATGCTGTGCTTACCCCTTTGAA 106
QY 61 TT 62
DB 107 GT 108
RESULT 7
AAH03775/C
ID AAH03775 standard; cDNA; 825 BP.
XX
XX AAH03775;
XX
XX 26-JUN-2001 (first entry)
DT
DE Human CDNA clone (5'-primer) SEQ ID NO:610.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length CDNA defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX CDNA.
XX
XX Claim 1; SEQ ID NO 610; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length CDNA defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification; where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence; where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX Sequence 825 BP; 281 A; 123 C; 163 G; 255 T; 0 U; 3 Other;

XX Query Match 50.9%; Score 35.6; DB 4; Length 825;

XX Best Local Similarity 71.2%; Pred. No. 0.0041; Mismatches 19; Indels 0; Gaps 0;

XX Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 4 TTGTTGCTGCTCTGCGAGCAGCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63
DB 458 TTTCCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTATTAACATTACCTTT 399
QY 64 CCAATT 69
DB 398 CTCCTT 393

RESULT 8

AAH17575/C
ID AAH17575 standard; cDNA; 1626 BP.

XX AAH17575;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:17065.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 17065; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dt primer and an oligonucleotide complementary to the

XX complementary strand of a polynucleotide which comprises one of the 5602

XX nucleotide sequences defined in the specification; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

XX Sequence 1626 BP; 509 A; 297 C; 299 G; 521 T; 0 U; 0 Other;

XX Query Match 50.9%; Score 35.6; DB 4; Length 1626;

XX Best Local Similarity 71.2%; Pred. No. 0.005; Mismatches 19; Indels 0; Gaps 0;

XX Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 4 TTGTTGCTGCTCTGCGAGCAGCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63
DB 913 TTTCCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTATTAACATTACCTTT 854
QY 64 CCAATT 69
DB 853 CTCCTT 848

RESULT 9

AAH17101/C
ID AAH17101 standard; cDNA; 2176 BP.

XX AAH17101;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:16427.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-

XX length cDNAs defined in the specification, and for the detection and/or

XX diagnosis of the abnormality of the proteins encoded by the full-length

XX cDNAs.

XX Claim 8; SEQ ID NO 16427; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-

XX length cDNAs defined in the specification. Where a primer set comprises:

XX (a) an oligo-dt primer and an oligonucleotide complementary to the


```

FT CDS 36..260
FT /*tag= a
FT /product= "CADECM-23"
XX
XX WO2004015396-A2.
XX
XX 19-FEB-2004.
XX
XX 12-AUG-2003; 2003WO-US025418.
XX
XX 13-AUG-2002; 2002US-0403781P.
XX 30-AUG-2002; 2002US-0407034P.
XX 13-SEP-2002; 2002US-0410566P.
XX 24-SEP-2002; 2002US-0413482P.
XX 25-SEP-2002; 2002US-0413890P.
XX 08-NOV-2002; 2002US-0424504P.
XX 13-NOV-2002; 2002US-0426222P.
XX
XX (INCY-) INCYTE CORP.
XX
XX ELLIOTT VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX Becha SD, Margulis JP, Swarnakar A, Chavla NK, Ramkumar J;
XX Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX Wang JT, Chien D, Yang YG;
XX
XX WPI; 2004-191795/18.
XX P-PSDB; ADM80794.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
XX diagnosing, treating and preventing immune, neurological, developmental,
XX connective tissue and cell proliferative disorders including cancer.
XX
XX Claim 5; SEQ ID NO 65; 272pp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
XX matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX cyostatic and anorectic activities. The CADECM polypeptides and
XX polynucleotides are useful in diagnosing, treating and preventing immune,
XX neurological, developmental, connective tissue and cell proliferative
XX disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX cancer, obesity and Tangier disease.
XX
XX Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;
SQ

```

```

Query Match 48.3%; Score 33.8; DB 12; Length 318;
Best Local Similarity 84.4%; Pred. No. 0.014;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 GATTTGTTGCTGCTGCTGAGAGAGCCTCAGTATGCCCTGT 45
Db 47 GATTTTATTGCTGCTGCTGAGAGAGCCTTTCAGTATGCCCTCT 91

```

Search completed: February 8, 2005, 13:57:04
Job time : 96.858 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 18.9057 Seconds
(Without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70

Sequence: 1 gattcttgctgctgctcc.....aaattccatttccattt 70

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	48.6	476	4 US-09-744-128-3	Sequence 3, App1
2	33.8	48.3	500	2 US-09-018-595B-3	Sequence 3, App1
3	33.8	48.3	500	3 US-09-324-709A-3	Sequence 3, App1
4	33.8	48.3	793	2 US-09-018-595B-1	Sequence 1, App1
5	33.8	48.3	793	3 US-09-324-709A-1	Sequence 1, App1
6	32.2	46.0	802	2 US-09-018-595B-2	Sequence 2, App1
7	32.2	46.0	802	3 US-09-324-709A-2	Sequence 2, App1
8	26.4	37.7	941	4 US-09-205-258-13	Sequence 13, App1
9	26.4	37.7	941	4 US-09-205-258-13	Sequence 13, App1
10	25.8	36.9	292	4 US-09-513-999C-32153	Sequence 32153, A
11	25	35.7	2871	4 US-09-328-352-134	Sequence 134, App
12	24.8	35.4	756	4 US-09-270-767-9938	Sequence 9938, Ap
13	24.8	35.4	1863	4 US-09-270-767-13233	Sequence 13233, A
14	24	34.3	592	3 US-08-896-164-80	Sequence 80, App1
15	23.6	33.7	742	1 US-08-266-080B-1	Sequence 1, App1
16	23.6	33.7	742	5 PCT-US95-05423-1	Sequence 1, App1
17	23.6	33.7	744	2 US-08-595-043A-76	Sequence 76, App1
18	23.6	33.7	952	4 US-09-636-368-2	Sequence 2, App1
19	23.6	33.7	1133	1 US-07-750-080A-2	Sequence 2, App1
20	23.6	33.7	1133	1 US-07-750-080A-3	Sequence 2, App1
21	23.6	33.7	1133	3 US-08-651-472-2	Sequence 2, App1
22	23.6	33.7	1133	3 US-08-651-472-2	Sequence 2, App1
23	23.6	33.7	1133	3 US-08-358-928-2	Sequence 2, App1
24	23.6	33.7	1133	3 US-08-358-928-3	Sequence 2, App1
25	23.6	33.7	1209	1 US-07-750-080A-22	Sequence 22, App1
26	23.6	33.7	1209	3 US-08-651-472-22	Sequence 22, App1
27	23.6	33.7	1209	3 US-08-358-928-22	Sequence 22, App1

28	23.6	33.7	3878	3 US-08-651-472-65	Sequence 65, App1
29	23.6	33.7	3878	3 US-08-358-928-65	Sequence 65, App1
30	23.6	33.7	4145	3 US-08-651-472-62	Sequence 62, App1
31	23.6	33.7	4145	3 US-08-358-928-62	Sequence 62, App1
32	23.6	33.7	4277	3 US-08-651-472-63	Sequence 63, App1
33	23.6	33.7	4277	3 US-08-358-928-63	Sequence 63, App1
34	23.6	33.7	4659	1 US-08-358-928-63	Sequence 10, App1
35	23.6	33.7	4701	3 US-08-651-472-64	Sequence 10, App1
36	23.6	33.7	4701	3 US-08-358-928-64	Sequence 64, App1
37	23.6	33.7	4818	1 US-08-232-463-11	Sequence 11, App1
38	23.6	33.7	4821	1 US-08-232-463-12	Sequence 12, App1
39	23.6	33.7	4824	3 US-08-232-463-13	Sequence 13, App1
40	23.6	33.7	5532	3 US-08-651-472-72	Sequence 72, App1
41	23.6	33.7	5532	3 US-08-358-928-72	Sequence 72, App1
42	23.6	33.7	6474	3 US-08-651-472-66	Sequence 66, App1
43	23.6	33.7	6474	3 US-08-358-928-66	Sequence 66, App1
44	23.6	33.7	6601	1 US-08-232-463-15	Sequence 15, App1
45	23.6	33.7	6710	4 US-09-355-040-28	Sequence 28, App1

ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27656/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3

Query Match      48.6%; Score 34; DB 4; Length 476;
Best Local Similarity 88.1%; Pred. No. 0.001;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GATTTGTTGCTGCTGCTGCTGAGACAGCTTCACTATGCC 42
Db      59 GATCTGTTGCTGCTGCTGCTGAGACAGCTTGTCTATGCC 100

RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Weitz,
; ADDRESSER: Wilson Somsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-018-595B-3

Query Match 48.3%; Score 33.8; DB 2; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.0013;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTGAGAGAGCTTCAGTANGCCTGT 45
Db 80 GATTATTGCTGCTGCTCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 3
US-09-324-709A-3
Sequence 3, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
TITLE OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-324-709A-3

Query Match 48.3%; Score 33.8; DB 3; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.0013;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTGAGAGAGCTTCAGTANGCCTGT 45
Db 80 GATTATTGCTGCTGCTCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

Query Match 48.3%; Score 33.8; DB 2; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.0015;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTGTGTTGCTGCTCTGAGAGAGCTTCAGTANGCCTGT 45
Db 80 GATTATTGCTGCTGCTCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 5
US-09-324-709A-1
Sequence 1, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSER: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-1

Query Match 48.3%; Score 33.8; DB 3; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.0015; 7; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGACGCTTCAGTATGCGCTGT 45
Db 80 GATTTATTTGCTGCTGCTGAGAGACGCTTTTGCAATGCGCTCT 124

RESULT 6
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSER: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2

Query Match 46.0%; Score 32.2; DB 2; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.0062; 8; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GATTTGTTGGCTGCTGCTGAGAGACGCTTCAGTATGCGCTGT 45
Db 80 GATTTGTTGGCTGCTGCTGAGAGACGCTTTTGCAATGCGCTCT 124

RESULT 7
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSER: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362

REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match 46.0%; Score 32.2; DB 3; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.0062;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATTGTGCTGCTGCTGCTGAGAGAGCTTCAGTATGCTGT 45
DB 80 GATTGTGCTGCTGCTGCTGAGAGAGCTTCAGTATGCTGT 124

RESULT 8
US-09-205-258-13/c
Sequence 13, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 13
LENGTH: 941
TYPE: DNA
ORGANISM: Homo sapiens
US-09-205-258-13

Query Match 37.7%; Score 26.4; DB 4; Length 941;
Best Local Similarity 69.2%; Pred. No. 1.2;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 17 CTCTGGAGAGAGCTTCAGTATGCTGTGAGTAAATTTCCATTTCCAT 68
DB 845 CTCTGGAGAGAGCTTCAGTATGCTGTGAGTAAATTTCCATTTCCAT 794

RESULT 9
US-09-205-258-218/c
Sequence 218, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06


```

/ EARLIER APPLICATION NUMBER: 60/048,880
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,896
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,020
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,876
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,895
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,884
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,894
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,971
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,882
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,899
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,893
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,900
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,901
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,892
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,915
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,019
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,970
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,972
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,916
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,373
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,875
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,374
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,917
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,949
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,974
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,883
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,897
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,898
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,962
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,963
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,877
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,878
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/070,923
/ EARLIER FILING DATE: 1997-12-18
/ EARLIER APPLICATION NUMBER: 60/092,921
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/094,657
/ EARLIER FILING DATE: 1998-07-30
/ NUMBER OF SEQ ID NOS: 1227

```

```

/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 218
/ LENGTH: 941
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-205-258-218

```

```

Query Match          37.7%; Score 26.4; DB 4; Length 941;
Best Local Similarity 69.2%; Pred. No. 1.2;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

QY      17  CTCCTGGAGACGCTTCAAGTATCCGCGTGAATAATTCCATTCCAA 68
DB      845  CTACTGGACACGCTTAATGCACTTGCATTAATTTCGTTCCAA 794

```

```

RESULT 10
US-09-513-999C-32153
/ Sequence 32153; Application US/09513999C
/ Patent No. 6783961
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Duclert, A.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
/ PATENT REFERENCE: 59,US2,REG
/ CURRENT APPLICATION NUMBER: US/09/513,999C
/ CURRENT FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/122,487
/ PRIOR FILING DATE: 1999-02-26
/ NUMBER OF SEQ ID NOS: 36681
/ SOFTWARE: Patent.pm
/ SEQ ID NO: 32153
/ LENGTH: 292
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-513-999C-32153

```

```

Query Match          36.9%; Score 25.8; DB 4; Length 292;
Best Local Similarity 60.9%; Pred. No. 1.3;
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

```

QY      2  ATTTGTTGCTGCTCTCTGGAGACGCTTGAATGCTGTGATAATTTCAA 61
DB      100  ATTTTATTTCTTATTTCTTATTTAGGAGAGTGTCAAAATGTTTCACTTAATTTCAGT 159
QY      62  TTCTCAATTT 70
DB      160  TTCTGTTT 168

```

```

RESULT 11
US-09-328-352-134
/ Sequence 134; Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO: 134
/ LENGTH: 2871
/ TYPE: DNA
/ ORGANISM: Acinetobacter baumannii
/ US-09-328-352-134

```

```

Query Match          35.7%; Score 25; DB 4; Length 2871;
Best Local Similarity 64.9%; Pred. No. 6.1;
Matches 37; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```

QY 14 TGGCTCTGGGAGACGCTTCAGTANGCCTGTGATTAATTTCCATTTCATTT 70
DB 737 TGGCGCAATATGACACCGATTATGAAAGCCAGATATTTATTTCCATATTCAAATTT 793

RESULT 12

US-09-270-767-9938/C
Sequence 9938, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9938
LENGTH: 756
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-9938

Query Match 35.4%; Score 24.8; DB 4; Length 756;
Best Local Similarity 63.3%; Pred. No. 4.4;
Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 2 ATTTTGTTCCTGCTCTCTGGAGACGCTTCAGTATGCTGTGATTAATTTCCAT 61
DB 181 ATTTGAAATCATCTACTGTGTGACGAAAGAGATTTGTTCTTCAGTAAATTTCCAT 122

RESULT 13

US-09-270-767-1323/C
Sequence 1323, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1323
LENGTH: 1863
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-09-270-767-1323

Query Match 35.4%; Score 24.8; DB 4; Length 1863;
Best Local Similarity 60.3%; Pred. No. 6.2;
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 3 TTTTGTTCCTGCTCTCTGGAGACGCTTCAGTATGCTGTGATTAATTTCCAT 62
DB 302 TTTTGTTCCTTTTACCCCGCTGCGGTCTGGCGGCTTGCTTTTGTTTTAAAT 243

QY 63 TCCAAATTT 70
DB 242 TCTCTCTT 235

RESULT 14

US-08-896-164-80/C
Sequence 80, Application US/08896164
Patent No. 6218521
GENERAL INFORMATION:
APPLICANT: OBATA, YUICHI
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED
TITLE OF INVENTION: WITH GASTRIC CANCER AND METHODS FOR
TITLE OF INVENTION: DIAGNOSING AND TREATING GASTRIC CANCER

NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/896,164
FILING DATE: July 17, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6218521man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5499 - JEL/NDH/SLH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 592 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-896-164-80

Query Match 34.3%; Score 24; DB 3; Length 592;
Best Local Similarity 59.1%; Pred. No. 8.2;
Matches 39; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 5 TTTTGTTCCTGCTCTCTGGAGACGCTTCAGTATGCTGTGATTAATTTCCATTT 64
DB 562 TTTGTTGACGCTTCAGGATTCATCTGTAGCGGTCTCAGCACTGCGGACGTG 503

QY 65 CAATTT 70
DB 502 TATTNT 497

RESULT 15

US-08-266-080B-1/C
Sequence 1, Application US/08266080B
Patent No. 5606031
GENERAL INFORMATION:
APPLICANT: Jack Lille
APPLICANT: Tadahiko Kohno
APPLICANT: Duane Bonam
TITLE OF INVENTION: Production of Biologically Active
TITLE OF INVENTION: Recombinant Neurotrophic Protein
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/266,080B
FILING DATE: 27-JUNE-1994
PRIOR APPLICATION DATA:

```

1      APPLICATION NUMBER: 08/240,122
2      FILING DATE: 09-MAY-1994
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: 08/087,912
5      FILING DATE: 06-JULY-1993
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: 07/680,681
8      FILING DATE: 04-APRIL-1991
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: 07/594,126
11     FILING DATE: 09-OCT-1990
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER: 07/547,750
14     FILING DATE: 02-JULY-1990
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER: 07/505,441
17     FILING DATE: 06-APRIL-1990
18     ATTORNEY/AGENT INFORMATION:
19     NAME: Barry J. Swanson
20     REGISTRATION NUMBER: 33,215
21     REFERENCE/DOCKET NUMBER: SYNE200CS
22     TELECOMMUNICATION INFORMATION:
23     TELEPHONE: (303) 793-3333
24     TELEFAX: (303) 793-3433
25     INFORMATION FOR SEQ ID NO: 1:
26     SEQUENCE CHARACTERISTICS:
27     LENGTH: 742 nucleotides
28     TYPE: nucleic acid
29     STRANDEDNESS: single
30     TOPOLOGY: linear
31     FEATURE:
32     NAME/KEY: nucleic acid sequence for human BDNF
33     US-08-266-0808-1

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 94.3248 Seconds
(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70

Sequence: 1 gattctgttcgtcgtcc.....aaattccattccattc 70

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCF_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCFUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	43.6	62.3	600	US-10-027-632-287049	Sequence 287049,
C 2	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 3	43.6	62.3	600	US-10-027-632-287049	Sequence 287049,
C 4	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 5	43.6	62.3	600	US-10-027-632-287050	Sequence 287050,
C 6	35.6	50.9	3037	US-10-723-860-1311	Sequence 1311, Ap
C 7	34	48.6	65	US-10-723-860-1311	Sequence 6937, Ap
C 8	33.8	48.3	500	US-10-104-774-3	Sequence 1137, Ap
C 9	33.8	48.3	500	US-10-455-150-3	Sequence 3, Appli
C 10	33.8	48.3	793	US-10-104-774-1	Sequence 1, Appli
C 11	33.8	48.3	793	US-10-455-150-1	Sequence 1, Appli
C 12	33.8	48.3	793	US-10-755-889-605	Sequence 605, App

13	32.2	46.0	802	13	US-10-104-774-2	Sequence 2, Appli
14	32.2	46.0	802	16	US-10-455-150-2	Sequence 2, Appli
15	32.2	46.0	802	18	US-10-723-860-5797	Sequence 5797, Ap
C 16	28.6	40.9	277	18	US-10-856-499-38	Sequence 38, Appl
C 17	27.8	39.7	1185	15	US-10-369-493-34889	Sequence 34889, A
C 18	27.2	38.9	129	17	US-10-437-963-53858	Sequence 53858, A
C 19	26.6	38.0	653122	13	US-10-087-192-226	Sequence 226, App
C 20	26.4	37.7	423	9	US-09-864-761-11109	Sequence 11109, A
C 21	26.4	37.7	567	9	US-09-864-761-16464	Sequence 16464, A
C 22	26.4	37.7	941	10	US-09-933-767-13	Sequence 13, Appl
C 23	26.4	37.7	941	10	US-09-933-767-218	Sequence 218, Appl
C 24	26.4	37.7	941	14	US-10-004-860-13	Sequence 13, Appl
C 25	26.4	37.7	941	14	US-10-004-860-218	Sequence 218, Appl
C 26	26.4	37.7	941	14	US-10-023-282-13	Sequence 13, Appl
C 27	26.4	37.7	941	14	US-10-023-282-218	Sequence 218, Appl
C 28	26.4	37.7	1658	13	US-10-027-632-264601	Sequence 264601, A
C 29	26.4	37.7	1658	13	US-10-027-632-264602	Sequence 264602, A
C 30	26.4	37.7	1658	15	US-10-027-632-264601	Sequence 264601, A
C 31	26.4	37.7	1658	15	US-10-027-632-264602	Sequence 264602, A
C 32	25.8	36.9	2079	13	US-10-027-632-97628	Sequence 97628, A
C 33	25.8	36.9	2079	13	US-10-027-632-100426	Sequence 100426, A
C 34	25.8	36.9	2079	13	US-10-027-632-100427	Sequence 100427, A
C 35	25.8	36.9	2079	15	US-10-027-632-97628	Sequence 97628, A
C 36	25.8	36.9	2079	15	US-10-027-632-100426	Sequence 100426, A
C 37	25.8	36.9	2079	15	US-10-027-632-100427	Sequence 100427, A
C 38	25.8	36.9	3537	15	US-10-369-493-45684	Sequence 45684, A
C 39	25.8	36.9	66973	13	US-10-087-192-574	Sequence 574, Appl
C 40	25.8	36.9	74665	18	US-10-719-993-6854	Sequence 6854, Ap
C 41	25.4	36.3	410	16	US-10-424-599-16385	Sequence 16385, A
C 42	25.2	36.0	1650	14	US-10-198-846-11735	Sequence 11735, A
C 43	25.2	36.0	48436	9	US-09-927-602-38	Sequence 38, Appl
C 44	25.2	36.0	96588	11	US-09-927-722-292	Sequence 292, Appl
C 45	25.2	35.7	661	13	US-10-027-632-273400	Sequence 273400, A

ALIGNMENTS

RESULT 1
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287049
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287049
Query Match
Best Local Similarity 62.3%; Score 43.6; DB 13; Length 600;
Pred. No. 3e-06;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58
DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

RESULT 2

US-10-027-632-287050/c
Sequence 287050, Application US/10027632
Publication No. US2002019837A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 1999-11-23

PRIOR FILING DATE: 1999-09-28

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287050

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287050

Query Match

Best Local Similarity 84.5%; Pred. No. 3e-06; Length 600;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58

DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

RESULT 3

US-10-027-632-287049/c
Sequence 287049, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 1999-11-23

PRIOR FILING DATE: 1999-09-28

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287050

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287050

Query Match

Best Local Similarity 84.5%; Pred. No. 3e-06; Length 600;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58

DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287049

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287049

Query Match

Best Local Similarity 84.5%; Pred. No. 3e-06; Length 600;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58

DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

RESULT 4

US-10-027-632-287050/c
Sequence 287050, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 1999-11-23

PRIOR FILING DATE: 1999-09-28

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287050

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287050

Query Match

Best Local Similarity 84.5%; Pred. No. 3e-06; Length 600;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58

DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

RESULT 5

US-10-723-860-1311
Sequence 1311, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:

APPLICANT: Aziz, Natscha

APPLICANT: Zlotnik, Albert

TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &

TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators

FILE REFERENCE: 05882.0193.NPUS01

CURRENT FILING DATE: 2003-11-26

US-10-723-860-1311

Query Match

Best Local Similarity 84.5%; Pred. No. 3e-06; Length 600;

Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTCTGCGAGAGAGCTTCAGTATGCTGTGAGTAAATTTC 58

DB 245 GATTTATTTGCTGCTGCTCTGCGAGAGAGCTTTGGCATGCTGTGAGTAAACACCC 188

;; PRIOR APPLICATION NUMBER: 60/429,739
;; PRIOR FILING DATE: 2002-11-26
;; NUMBER OF SEQ ID NOS: 8393
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 1311
;; LENGTH: 259202
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 62.3%; Score 43.6; DB 18; Length 259202;
Best Local Similarity 84.5%; Pred. No. 1.8e-05;
Matches 49; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAGTATGCTGTGATAAATTTCC 58
Db 68884 GATTTATTTGCTGCTGCTGAGAGAGCTTTTGCATGCTGTGATAAACCACC 69941

RESULT 6
US-10-723-860-6937/C
Sequence 6937, Application US/10723860
Publication No. US20040253606A1

GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Gansburg, Wendy M.
APPLICANT: Ziornik, Albert
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6937
LENGTH: 3037
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1804)..(1823)
OTHER INFORMATION: n 1a a, c, g, or t
US-10-723-860-6937

Query Match 50.9%; Score 35.6; DB 18; Length 3037;
Best Local Similarity 71.2%; Pred. No. 0.0053;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 4 TTTGTTGCTGCTGCTGAGAGAGCTTCAGTATGCTGTGATAAATTTCCATTT 63
Db 913 TTTCTTGACTGCTGCTGAGAGAGCTTAATTCAGTTTCTGTTATTAACATTTACCTT 854

Qy 64 CCATTT 69
Db 853 CTCCTT 848

RESULT 7
US-09-908-975-1137

Sequence 1137, Application US/09908975
Publication No. US20030165843A1
GENERAL INFORMATION:
APPLICANT: SHOSHAN, Avi
APPLICANT: WASSERMAN, Alon
APPLICANT: MINTZ, Eli
APPLICANT: MINTZ, Liat
APPLICANT: FAIGER, Shimon
TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
FILE REFERENCE: 36688-0005
CURRENT APPLICATION NUMBER: US/09/908,975

;; CURRENT FILING DATE: 2001-07-20
;; PRIOR APPLICATION NUMBER: US 60/287,724
;; PRIOR FILING DATE: 2001-05-02
;; PRIOR APPLICATION NUMBER: US 60/221,607
;; PRIOR FILING DATE: 2000-07-28
;; NUMBER OF SEQ ID NOS: 32337
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 1137
;; LENGTH: 65
;; TYPE: DNA
;; ORGANISM: Rattus norvegicus
US-09-908-975-1137

Query Match 48.6%; Score 34; DB 10; Length 65;
Best Local Similarity 88.1%; Pred. No. 0.0069;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGAGCTTCAGTATGCC 42
Db 6 GATCTGTTGCTGCTGCTGAGAGAGAGCTTTGCTATGCC 47

RESULT 8
US-10-104-774-3
Sequence 3, Application US/10104774
Publication No. US20020164630A1

GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE .
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 500
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-3

Query Match 48.3%; Score 33.8; DB 13; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.015;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGAGCTTCAGTATGCCCTGT 45
Db 80 GATTTATTTGCTGCTGCTGAGAGAGAGCTTTGCCATGCCCTT 124

RESULT 9
US-10-455-150-3

Sequence 3, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSER: David J. Weitz,
Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible

```
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match      48.3%; Score 33.8; DB 16; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.015; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCCTCGGAGAGACGCTTCAGTATGCTGT 45
DB 80 GATTATTGCTGCTCCTCGGAGAGACGCTTTGCCATGCTCT 124

RESULT 10
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1

Query Match      48.3%; Score 33.8; DB 13; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCCTCGGAGAGACGCTTCAGTATGCTGT 45
DB 80 GATTATTGCTGCTCCTCGGAGAGACGCTTTGCCATGCTCT 124

RESULT 11
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
```

```
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSER: David J. Weitz,
Wilson Sonaini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match      48.3%; Score 33.8; DB 16; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCCTCGGAGAGACGCTTCAGTATGCTGT 45
DB 80 GATTATTGCTGCTCCTCGGAGAGACGCTTTGCCATGCTCT 124

RESULT 12
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 605
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-605

Query Match      48.3%; Score 33.8; DB 17; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.017; Indels 7; Gaps 0;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```


Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCTGT 45
|||
DB 80 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 13
US-10-104-774-2
; Sequence 2, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; PRIOR FILING DATE: 2002-03-21
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 802
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-2

Query Match 46.0%; Score 32.2; DB 13; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.07;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCTGT 45
|||
DB 80 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 14
US-10-455-150-2
; Sequence 2, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-455-150-2

Query Match 46.0%; Score 32.2; DB 16; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.07;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCTGT 45
|||
DB 80 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCGATGCTCT 124

RESULT 15
US-10-723-860-5797
; Sequence 5797, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5797
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match 46.0%; Score 32.2; DB 18; Length 852;
Best Local Similarity 82.2%; Pred. No. 0.071;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCTGT 45
|||
DB 80 GATTTATTTGCTGCTGCTGAGAGAGCTTTGCGATGCTCT 124

Search completed: February 9, 2005, 06:51:01
Job time : 97.3248 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 833.766 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-7

Perfect score: 70
Sequence: 1 gatttcgttcgttcgttcgcctcc.....aaattccaatccattc 70

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	55.1	439	6	CB473334 sn68_E03.
2	35.6	50.9	303	7	W33932 md54d02.r1
3	35.6	50.9	430	7	W40649 mc43h12.r1
4	35.6	50.9	441	1	A1798381 tr34c02.x
5	35.6	50.9	444	2	BE220076 hv6h03.x
6	35.6	50.9	448	1	A1338491 q997h05.x
7	35.6	50.9	536	6	W36345 mb72c12.r1
8	35.6	50.9	561	5	BX470683 DKZP686N
9	35.6	50.9	585	5	AY419441 Mus muscu
10	35.6	50.9	586	2	BF001849 7994c09.x
11	35.6	50.9	588	4	BF448291 nad1f04.
12	35.6	50.9	620	4	BG622884 602647718
13	35.6	50.9	645	6	CD773419 AGENCCOURT
14	35.6	50.9	707	6	CB056709 NISC_j119
15	35.6	50.9	707	7	CN360063 170005830
16	35.6	50.9	825	3	AK029358 Mus muscu
17	35.6	50.9	825	1	AU118731 AU118731
18	35.6	50.9	843	6	CB588212 AGENCCOURT
19	35.6	50.9	859	6	CB588525 AGENCCOURT
20	35.6	50.9	865	6	CB589177 AGENCCOURT
21	35.6	50.9	886	6	CB587051 AGENCCOURT
22	35.6	50.9	889	6	CB589251 AGENCCOURT
23	35.6	50.9	891	6	CB574837 AGENCCOURT
24	35.6	50.9	913	6	CB590451 AGENCCOURT

25	35.6	50.9	919	6	CB587332 AGENCCOURT
26	35.6	50.9	928	6	CB590111 AGENCCOURT
27	35.6	50.9	945	2	BB614068 BB614068
28	35.6	50.9	990	7	W12906 W12906
29	35.6	50.9	1006	7	W29475 W29475
30	35.6	50.9	1020	7	W08102 W08102
31	34.6	49.4	293	7	R47143 R47143
32	34.6	49.4	566	7	R46934 R46934
33	34.6	48.6	217	7	R46913 R46913
34	34.6	48.6	223	2	BB571643 BB571643
35	34.6	48.6	395	7	R46955 R46955
36	34.6	48.6	449	7	R47024 R47024
37	34.6	48.6	1031	6	CB575508 CB575508
38	33.8	48.3	570	9	AY419439 AY419439
39	33.8	48.3	570	9	AY419440 AY419440
40	33.6	48.0	303	7	R46903 R46903
41	33.6	48.0	318	7	R47078 R47078
42	33.6	48.0	467	7	R46947 R46947
43	33.6	47.1	373	7	R47135 R47135
44	33.6	47.1	395	7	R47030 R47030
45	32.4	46.3	257	7	R47100 R47100

ALIGNMENTS

RESULT 1
CB473334/c 429 bp mRNA linear EST 26-MAR-2003
LOCUS CB473334
DEFINITION sn68_E03.f sn Sus scrofa CDNA 5', mRNA sequence.
ACCESSION CB473334
VERSION CB473334.1 GI:29279720
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE
AUTHORS Neilan,J.G., Kutish,G.P., Lu,Z., Zaak,A. and Rock,D.L.
TITLE Sequence analysis of African swine fever virus infected and non-infected porcine macrophage CDNA libraries
JOURNAL Unpublished (2003)
COMMENT Contact: Neilan JG

Plum Island Animal Disease Center
US Department of Agriculture, Agricultural Research Service
PO Box 848, Greenport, NY 11944-848, USA
Tel: 631 323 3133
Fax: 631 323 3044

Email: jneil@pladac.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified by cross_match v0.990329 and Lucy v1.17p.
Seq primer: M13 Forward.

FEATURES
source Location/Qualifiers
1..429

/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/issue_type="lymphoid"
/cell_type="macrophage"
/lab_host="DH10B"

/clone_lib="sn"
/note="Vector: pSPORT1, Site 1: NotI, Site 2: SalI;
Library made from pools of polya selected RNA, isolated at different times post-infection (0 to 16 hrs) from African swine fever virus (isolate Pretoriuskop/96/4) infected swine macrophages. Macrophages were derived from peripheral blood mononuclear cells cultured for 48 hrs on plastic in the presence of 30% U937 supernatant."

ORIGIN
Query Match 55.1%; Score 38.6; DB 6; Length 429;
Best Local Similarity 91.1%; Pred. No. 0.013;

Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAATGCTGT 45
 |||||
 Db 357 GATTTGTTGCTGCTGCTGAGAGAGCTTCTATGCTCT 313

RESULT 2
 W33932 303 bp mRNA linear EST 11-SEP-1996
 LOCUS mb5402.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
 DEFINITION IMAGE:333219.5 similar to gb:M10095 Mouse amelogenin (MOUSE);

ACCESSION W33932
 VERSION W33932.1 GI:1315837
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

REFERENCE
 AUTHORS Bukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 303)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:214619
 Seq primer: mob.REGA+RT
 High quality sequence stop: 185.
 Location/Qualifiers

FEATURES
 source
 1..303
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:333219"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NMF19.5"
 /note="Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 303;
 Best Local Similarity 90.5%; Pred. No. 0.13;
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAATGCTGT 42
 |||||
 Db 81 GATTTGTTGCTGCTGCTGAGAGAGCTTCTATGCTCT 122

RESULT 3
 W40649

LOCUS W40649 430 bp mRNA linear EST 11-SEP-1996
 DEFINITION mc43h12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
 IMAGE:351335.5 similar to gb:M10095 Mouse amelogenin (MOUSE);

ACCESSION W40649
 VERSION W40649.1 GI:1324982
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

REFERENCE
 AUTHORS Bukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:223135
 Seq primer: ERTPrimer
 High quality sequence stop: 323.
 Location/Qualifiers

FEATURES
 source

1..430
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:351335"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NMF19.5"
 /note="Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 430;
 Best Local Similarity 90.5%; Pred. No. 0.15;
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTGCTGAGAGAGCTTCAATGCTGT 42
 |||||
 Db 69 GATTTGTTGCTGCTGCTGAGAGAGCTTCTATGCTCT 110

RESULT 4

LOCUS A1798381 441 bp mRNA linear EST 06-JUN-1999
 DEFINITION tr34c02.x1 NCI CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2220194.3';
 mRNA sequence.

ACCESSION A1798381
 VERSION A1798381.1 GI:5363853
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 441)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdnp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers

FEATURES
source

1..441
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2220194"
/issue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/clone_1lb="NCI-CGAP_Ov23"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"

ORIGIN

Query Match 50.9%; Score 35.6; DB 1; Length 441;
Best Local Similarity 71.2%; Pred. No. 0.15;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTCTGAGACAGCTTCAGTACCTGTGAGTAAATTTCCATT 63
|||||
DB 69 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTACCTT 128
|||||
QY 64 CCAATT 69
|||||
DB 129 CTCCTT 134

RESULT 5
LOCUS BE220076 444 bp mRNA linear EST 03-JUL-2000
DEFINITION hv66h03.x1 NCI-CGAP_Lu24 Homo sapiens CDNA clone IMAGE:3178421 3',
mRNA sequence.
ACCESSION BE220076
VERSION BE220076
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 444)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be

FEATURES

source

found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 388.
Location/Qualifiers

1..444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3178421"
/issue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pTZ19-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI-CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (cloneIDB
1414920-1417991 and 1520904-1522439). Subtraction by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 50.9%; Score 35.6; DB 2; Length 444;
Best Local Similarity 71.2%; Pred. No. 0.15;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTCTGAGACAGCTTCAGTACCTGTGAGTAAATTTCCATT 63
|||||
DB 65 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTACCTT 124
|||||
QY 64 CCAATT 69
|||||
DB 125 CTCCTT 130

RESULT 6
LOCUS AI338491 448 bp mRNA linear EST 13-FEB-1999
DEFINITION q997h05.x1 Soares fetal fetus Nb2HF8_9w Homo sapiens CDNA clone
IMAGE:193953 3', mRNA sequence.
ACCESSION AI338491
VERSION AI338491
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 448)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1345 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 447.
Location/Qualifiers

FEATURES

source

1..448
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:193953"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/clone_1lb="Soares fetal fetus Nb2HF8_9w"
/note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo (dT) primer [5' TGTTACCAATCTAGATGGAGGCGCGCTTAATTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 50.9%; Score 35.6; DB 1; Length 448;
Best Local Similarity 71.2%; Pred. No. 0.15;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGGCTGCTGCTGAGAGCGCTTACAGTACCCGTAGTAATTTCCAAATT 63
|||||
DB 69 TTTCTTGACTGCTGCTGAGAGCTAAATTCAGTTTCTGTATTACATTAACCTTT 128
|||||
QY 64 CCAATT 69
|||
DB 129 CTCCTT 134

RESULT 7
W36345 536 bp mRNA linear EST 11-SEP-1996
LOCUS M36345
DEFINITION MB72C12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:334966 5' similar to gp:M10095 Mouse amelogenin (MOUSE);
mRNA sequence.
W36345 GI:1318120
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 536)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Gaisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, B.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Tietelberg, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marras M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbl.gov) for further information.
MGI:216366
Seq primer: ETPprimer
High quality sequence stop: 359.

FEATURES

source
1..536
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:334966"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_1ib="Soares mouse p3NMF19.5"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTAGATGGAGGCGCGCTTAATTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento

ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 536;
Best Local Similarity 90.5%; Pred. No. 0.15;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTTGGTGGCTGCTGCTGAGAGCGCTTACAGTACGCC 42
|||||
DB 64 GATTTTGGTGGCTGCTGCTGAGAGCGCTTGTGCTAGCC 105
|||||

RESULT 8
BX470683/c 561 bp mRNA linear EST 04-SEP-2003
LOCUS BX470683
DEFINITION DKFZp686N21110.r1 686 (synonym: h1cc3) Homo sapiens cDNA clone
DKFZp686N21110 5', mRNA sequence.
BX470683
ACCESSION BX470683
VERSION BX470683.1 GI:31665009
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 561)
Koehrer, K., Beyer, A., Mewes, H.W., Well, B., Amid, C., Osanger, A.,
Fodor, G., Han, W., and Wiemann, S.
EST (Koehrer, K., Beyer, A., Mewes, H.W., Well, B., Amid, C., et al.)
Unpublished (2003)
Contact: MIPS

TITLE Ingolstaedter Landstr.1, D-85764 Neuberg, Germany
JOURNAL This is the 5' sequence of the clone insert.
COMMENT This is the 5' sequence of the clone insert.
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Heinrich-
Heine-University, Dusseldorf/Germany) within the CDNA sequencing
consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp686N21110) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
1..561
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686N21110"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1ib="686 (synonym: h1cc3)"
/note="Vector: pT73D; Site 1: SfiI; Site 2: SfiI;
cDNA-collection"

ORIGIN

Query Match 50.9%; Score 35.6; DB 5; Length 561;
Best Local Similarity 71.2%; Pred. No. 0.15;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTGGCTGCTGCTGAGAGCGCTTACAGTACCCGTAGTAATTTCCAAATT 63
|||||
DB 211 TTTCTTGACTGCTGCTGAGAGCTAAATTCAGTTTCTGTATTACATTAACCTTT 152
|||||
QY 64 CCAATT 69
|||
DB 151 CTCCTT 146

RESULT 9
AY419441 585 bp DNA linear GSS 17-DEC-2003
LOCUS AY419441
DEFINITION Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.

ACCESSION AY419441
 VERSION AY419441.1 GI:39775398
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 585)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferriere,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PubMed 14671302
 REFERENCE 2 (bases 1 to 585)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferriere,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J., Adams,M.D. and Cargill,M.
 TITLE Direct Submision
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.
 FEATURES
 source Location/Qualifiers
 1..585
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>585
 /gene="AMELX"
 /locus_tag="HMC6891"
 ORIGIN
 Query Match 50.9%; Score 35.6; DB 9; Length 585;
 Best Local Similarity 90.5%; Pred. No. 0.16;
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 42
 12 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 53
 Db 12 GATTTTGTTCCTGCTGCTGAGAGAGCCTTCAGTATGCC 53
 RESULT 10
 BFO01849 586 bp mRNA linear EST 06-OCT-2000
 LOCUS 7994c09.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:3314128 3',
 DEFINITION mRNA sequence.
 ACCESSION BFO01849
 VERSION BFO01849.1 GI:10702124
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 586)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seg primer: -40UP from Gibco

FEATURES
 source High quality sequence stop: 471.
 Location/Qualifiers
 1..586
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3314128"
 /tissue_type="colon tumor, RER+"
 /lab_host="DH10B"
 /lab_host="NCI CGAP Col6"
 /note="Organ: colon; Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalised library NCI CGAP Col6 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1057416-1061255, and 114584-1145351).
 Subtraction by Bento Soares and M. Fatima Bonaldo. "
 ORIGIN
 Query Match 50.9%; Score 35.6; DB 2; Length 586;
 Best Local Similarity 71.2%; Pred. No. 0.16;
 Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 QY 4 TTTGTTGCTGCTGCTGAGAGAGCCTTCAGTATGCCATTTCAATT 63
 65 TTTGTTGCTGCTGCTGAGAGAGCCTTCAGTATGCCATTTCAATT 124
 Db 65 TTTGTTGCTGCTGCTGAGAGAGCCTTCAGTATGCCATTTCAATT 124
 QY 64 CCAATT 69
 125 CTCCTT 130
 Db 125 CTCCTT 130
 RESULT 11
 BFA48291 588 bp mRNA linear EST 01-DEC-2000
 LOCUS nad17f04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:336587 3',
 DEFINITION mRNA sequence.
 ACCESSION BFA48291
 VERSION BFA48291.1 GI:11513814
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 588)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seg primer: -40UP from Gibco
 High quality sequence stop: 465.
 Location/Qualifiers
 1..588
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:336587"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /clone_id="NCI CGAP Lu24"
 /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a

ORIGIN

modified polylinker; plasmid DNA from the normalized library NCI CGAP-Lus was prepared, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

Query Match 50.9%; Score 35.6; DB 2; Length 588;

Best Local Similarity 71.2%; Pred. No. 0.16; Mismatches 19; Indels 0; Gaps 0;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTCTGGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63
 DB 69 TTTCCTTGACTGCTCATGAGACTTAATTCAGTTTCTGTATTAACTTACACTTT 128
 QY 64 CCAATT 69
 DB 129 CTCCTT 134

RESULT 12
 BG622884/c 620 bp mRNA linear EST 18-APR-2001
 LOCUS 602647718F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4769213 5',
 DEFINITION mRNA sequence.

ACCESSION BG622884.1 GI:13674255

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 620)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strauberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LNCM634 row: c column: 06

High quality sequence stop: 618.

Location/Qualifiers

1. 620

/organism="Homo sapiens"

/mol_type="RNA"

/db_xref="taxon:9606"

/clone="IMAGE:4769213"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: placenta; Vector: PDNR-LIB (Clontech);

Site 1: SfiI (ggcgctcggc); Site 2: SfiI

(ggcgctcggc); 5' and 3' adaptors were used in cloning

as follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3'

and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGCGCATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.3

kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 50.9%; Score 35.6; DB 4; Length 620;

Best Local Similarity 71.2%; Pred. No. 0.16;

Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTCTGGAGACCTTCAGTATGCTGTGAGTAAATTTCCAAATT 63

DB 513 TTTCCTTGACTGCTCATGAGACTTAATTCAGTTTCTGTATTAACTTACACTTT 454

QY 64 CCAATT 69

DB 453 CTCCTT 448

RESULT 13

CD773419 645 bp mRNA linear EST 02-JUL-2003

LOCUS AGENCOURT.14713373 NIH_MGC_190 Mus musculus cDNA clone

IMAGE:30501267 5', mRNA sequence.

ACCESSION CD773419

VERSION CD773419.1 GI:32431921

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 645)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Yoshiniko Yamada, Takashi Nakamura, NIDCR

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: NDCM223 row: k column: 04

High quality sequence stop: 611.

Location/Qualifiers

1. 645

/organism="Mus musculus"

/mol_type="RNA"

/db_xref="taxon:10090"

/clone="IMAGE:30501267"

/lab_host="DH10B (phage-resistant)"

/note="Organ: Pooled - Molax; Vector: PDNR-LIB; Site 1:

SfiI (ggcgctcggc); Site 2: SfiI (ggcgctcggc);

Non-normalized full-length enriched library 5' and 3'

adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGGCGCGCATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.71

kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by

PCR. This library was enriched for full-length clones and

was constructed by Clontech Laboratories (Palo Alto, CA)

Corp."

ORIGIN

Query Match 50.9%; Score 35.6; DB 6; Length 645;

Best Local Similarity 90.5%; Pred. No. 0.16; Mismatches 4; Indels 0; Gaps 0;

Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCC 42

DB 84 GATTTGTTGCTGCTCTCTGGAGACCTTCAGTATGCC 125

RESULT 14

CB056709

LOCUS CB056709 707 bp mRNA linear EST 17-JAN-2003
 DEFINITION NISC J119905 w1 Soares NMBP13-15 Mus musculus cDNA clone
 IMAGE:4848584 5', mRNA sequence.
 ACCESSION CB056709
 VERSION CB056709.1 GI:27794996
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 707)
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Straubeberg, Ph.D.
 Email: CGAPbs-remail.nih.gov
 CDNA Library Preparation:
 DNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
 Sequencing Center (NISC)
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
infoimage.lnl.gov
 Plate: L1AM10779 row: N column: 9
 Seq primer: T7 primer
 Location/Qualifiers
 1..707
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:4848584"
 /tissue_type="pituitary gland"
 /dev_stage="juvenile, 13-15 days"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="Soares NMBP13-15"
 /note="Organ: brain; Vector: pTZ19-Pac1; Site_1: NotI;
 Site_2: EcoRI; 1st strand cDNA was primed with a NotI-
 oligo(dT) primer
 5'-ACCTGAGAGATTGCGCGCGCTGACCAATGTTTCTTTTCTTTT-3';
 double-stranded cDNA was ligated to EcoRI adaptors
 5'-ATTGCGACAGAG-3' and 5'-CTCGTCCG-3' (Pharmacia),
 digested with NotI and cloned into the NotI and EcoRI
 sites of the pTZ19-Pac1 vector. Library went through one
 round of normalization, and was constructed in the
 laboratory of W. Bento Soares (University of Iowa)."

ORIGIN

Query Match 50.9%; Score 35.6; DB 6; Length 707;
 Best Local Similarity 90.5%; Pred. No. 0.16;
 Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTTCTGCTGCTGAGAGAGCCCTTCAATGCC 42
 |||||
 DB 58 GATTTGTTTCTGCTGCTGAGAGAGCCCTTTCATGCC 99
 |||||

RESULT 15
 CN360063 707 bp mRNA linear EST 16-MAY-2004
 LOCUS 17000583090269 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
 DEFINITION CN360063
 ACCESSION CN360063
 VERSION CN360063.1 GI:47359997
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 707)
 REFERENCE Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murae, J., Fisk, G.J.,
 Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
 Lebkowksi, J. and Scanlon, L.W.
 TITLE Transcriptome characterization elucidates signaling networks that

JOURNAL control human ES cell growth and differentiation
 Nat. Biotechnol. 22 (6), 707-716 (2004)
 COMMENT Contact: Brandenberger R
 Regenerative Medicine
 Genon Corporation
 230 Constitution Drive, Menlo Park, CA 94025, USA
 Tel: 650 473 8658
 Fax: 650 473 7760
 Email: rbrandenberger@genon.com
 Insert Length: 707 Std Error: 0.00.
 Location/Qualifiers
 1..707
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cells, DMSO-treated H9 cell
 line"
 /clone_1lb="GRN PREHEP"
 /note="oligo dt primed, full-length enriched cDNA library
 from DMSO-treated hES cell line H9 (p22) maintained in
 feeder-free conditions".

ORIGIN

Query Match 50.9%; Score 35.6; DB 7; Length 707;
 Best Local Similarity 71.2%; Pred. No. 0.16;
 Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 TTTGTTTCTGCTGCTGAGAGAGCCCTTCAATGATGCTGTAGTAAATTTCAATTT 63
 |||||
 DB 596 TTTCTTGAAGCTGCTGATGAGAGCTAAATTCAGTTTCTGTATTACATTTACACTTT 537
 |||||

QY 64 CCAATT 69
 |||||
 DB 536 CTCCTT 531
 |||||

Search completed: February 8, 2005, 22:18:08
 Job time : 835.766 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 373.57 Seconds

(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67
Sequence: 1 ccaacagctgcgaataaag.....gtttaagacagtcacagacc 67

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Genbankl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srb:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.8	71.3	5712	4	AB091793 Equus cab
2	47.8	71.3	6451	4	AB091789 Bos tauru
3	44.6	66.6	5684	9	AB091785 Lemur cat
4	39.8	59.4	5562	9	AB091787 Oclemur
5	37.2	55.5	6931	4	AB091792 Sus scrof
6	37	55.2	6465	9	AB091781 Pan trogl
7	36.6	54.6	7425	4	AB091791 Sus scrof
8	36.2	54.0	2501	9	D83729 Homo sapien
9	36.2	54.0	177654	2	AP000918 Homo sapi
10	36.2	54.0	200214	9	AC013412 Homo sapi
11	35.4	52.8	1935	9	D83730 Homo sapien
12	35	52.2	5591	4	AB091794 Equus cab
13	34.8	51.9	6442	9	AB091783 Saimiri s
14	34.6	51.6	7163	9	AB091782 Pan trogl
15	34.6	51.6	38785	9	BS000568 Pan trogl
16	34.6	51.6	177726	9	AC145770 Pan trogl
17	34.6	51.6	190089	9	BS000576 Pan trogl
18	33.8	50.4	8810	9	AY040206 Homo sapi
19	33.8	50.4	158142	2	AL357130 Homo sapi

20	33.8	50.4	259202	9	AC002366	AC002366 Human Xp2
21	33.2	49.6	6585	11	BV089295	BV089295 RPAMSE00
22	33.2	49.6	695	11	BV097603	BV097603 RPAMSE00
23	31.6	47.2	7454	9	AB091784	AB091784 Saimiri s
24	31	46.3	463	10	D8306382	D83064 Mus muscu
25	31	46.3	9384	10	AF294397	AF294397 Mus muscu
26	31	46.3	95826	10	AL805974	AL805974 Mouse DNA
27	28.4	42.4	168200	2	AC120631	AC120631 Rattus no
28	28.4	42.4	235978	2	AC094209	AC094209 Rattus no
29	28.2	42.1	135141	10	BX005253	BX005253 Mouse DNA
30	28.2	42.1	244095	2	AC131841	AC131841 Rattus no
31	27.8	41.5	169816	2	AC118600	AC118600 Rattus no
32	27.8	41.5	186329	2	AC107238	AC107238 Mus muscu
33	27.8	41.5	236275	10	AC105297	AC105297 Mus muscu
34	27.8	41.5	242603	2	AC121056	AC121056 Rattus no
35	27.6	41.2	150635	4	AC130186	AC130186 Ateles
36	27.6	41.2	155547	10	AC132595	AC132595 Mus muscu
37	27.6	41.2	203046	2	AC139958	AC139958 Rattus no
38	27.6	41.2	206765	2	AC148973	AC148973 Mus muscu
39	27.6	41.2	234248	2	AC113775	AC113775 Rattus no
40	27.6	41.2	246315	2	AC094552	AC094552 Rattus no
41	27.6	41.2	249661	2	AC094299	AC094299 Rattus no
42	27.2	40.6	5668	9	BSM805696	BX537450 Homo sapi
43	27.2	40.6	78347	2	AC022642	AC022642 Homo sapi
44	27.2	40.6	118572	9	AL133408	AL133408 Human DNA
45	27.2	40.6	170795	2	AC021346	AC021346 Homo sapi

ALIGNMENTS

RESULT 1
AB091793
LOCUS AB091793 5712 bp DNA linear MAM 02-MAY-2003
DEFINITION Equus caballus AMELX gene for amelogenin, partial cds.
ACCESSION AB091793
VERSION AB091793.1 GI:29126038

KEYWORDS
SOURCE Equus caballus (horse)
ORGANISM

REFERENCE
1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

JOURNAL
MEDLINE
PUBMED
12672962
2 (bases 1 to 5712)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submision
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokenai), Department of Biosystems Science;
Shonan Kikusaijima, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.wol.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES

FEATURES
source
1. 5712
Location/Qualifiers

gene
5'UTR
CDS
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/sex="male"
174. .5712
/gene="AMELX"
/db_xref="taxon:9796"
/product="amelogenin"
/protein_id="BAC66112.1"
/db_xref="GI:29126039"

ORIGIN

/translation="MGTWILFASLGAAPALPLPSPHGPCTINSYEVLTPTKXOS
LIRQPTSYGYEPMGMLHRIIPLVLSQNPNSMALQPHHILPMVSAQHPVVPQCPMM
PLPGHSMVPTQHHQNPPLRPVQCPHPPVOPPOPPPOPPPOPPPLHPIQPLPPOPLP
PIFPIQPLPMLPDLPLEAMPATRG"

Query Match 71.3%; Score 47.8; DB 4; Length 5712;
Best Local Similarity 82.1%; Pred. No. 8e-07;
Matches 55; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTGTACTTAAAGGTTAAGACAGTA 60
|||||
Db 1669 CACAAGCTTGAATAAATCTCTCCACGGTTGTAACCTTAAGGTTAAGACAGTA 1728
|||||

Qy 61 CAAGATC 67
|||||
Db 1729 CAAGATC 1735

RESULT 2
AB091789 6451 bp DNA linear MAM 02-MAY-2003

LOCUS Bos taurus AMELX gene for amelogenin, partial cds.
AB091789
ACCESSION
VERSION AB091789.1 GI:29126030
KEYWORDS

SOURCE
ORGANISM Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

REFERENCE
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569

FEATURES
PUBMED 12672962
2 (bases 1 to 6451)
REFERENCE Iwase, M., Satta, Y. and Takahata, N.
AUTHORS Direct Submission
TITLE Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryuwo1.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
LOCATION/Qualifiers

FEATURES
SOURCE 1. .6451
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/sex="male"
238. .6451
/gene="AMELX"
5'UTR join(238. .340,1650. .1661)
/gene="AMELX"
CDS join(1662. .1715,4173. .4220,5560. .5601,5665. .5709,
5987. .56451)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66108.1"
/db_xref="GI:29126031"

/translation="MGTWILFASLGAAPALPLPSPHGPCTINSYEVLTPTKXOS
LIRQPTSYGYEPMGMLHRIIPLVLSQNPNSMALQPHHILPMVSAQHPVVPQCPMM
PLPGHSMVPTQHHQNPPLRPVQCPHPPVOPPOPPPOPPPOPPPLHPIQPLPPOPLP
PIFPIQPLPMLPDLPLEAMPATRG"

ORIGIN

Query Match 71.3%; Score 47.8; DB 4; Length 6451;
Best Local Similarity 82.1%; Pred. No. 8e-07;
Matches 55; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTGTACTTAAAGGTTAAGACAGTA 60
|||||
Db 1755 CACAAGCTTGAATAAAGTCTGCCACCAGTGTACTTAAAGGTTAAGACAGTA 1814
|||||

Qy 61 CAAGATC 67
|||||
Db 1815 CATGATC 1821

RESULT 3
AB091785 5684 bp DNA linear PRI 02-MAY-2003

LOCUS Lemur catta AMELX gene for amelogenin, partial cds.
AB091785
ACCESSION
VERSION AB091785.1 GI:29126023
KEYWORDS

SOURCE
ORGANISM Lemur catta (ring-tailed lemur)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsithini; Lemniridae; Lemur.

REFERENCE
AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569

FEATURES
PUBMED 12672962
2 (bases 1 to 5684)
REFERENCE Iwase, M., Satta, Y. and Takahata, N.
AUTHORS Direct Submission
TITLE Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryuwo1.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
LOCATION/Qualifiers

FEATURES
SOURCE 1. .5684
/organism="Lemur catta"
/mol_type="genomic DNA"
/db_xref="taxon:9447"
/sex="male"
100. .5684
/gene="AMELX"
5'UTR join(100. .156,1503. .1514)
/gene="AMELX"
CDS join(1515. .1568,3465. .3512,4804. .4845,4937. .4981,
5229. .5684)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66105.1"
/db_xref="GI:29126024"

/translation="MGTWILFASLGAAPALPLPSPHGPCTINSYEVLTPTKXOS
LIRQPTSYGYEPMGMLHRIIPLVLSQNPNSMALQPHHILPMVSAQHPVVPQCPMM
PLPGHSMVPTQHHQNPPLRPVQCPHPPVOPPOPPPOPPPOPPPLHPIQPLPPOPLP
PIFPIQPLPMLPDLPLEAMPATRG"

ORIGIN

Query Match 66.6%; Score 44.6; DB 9; Length 5684;
Best Local Similarity 79.1%; Pred. No. 1.2e-05;
Matches 53; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 CACAAGCTTGAATAAAGTCTGCCACCAGTGTACTTAAAGGTTAAGACAGTA 60
|||||
Db 1608 CACAAGCTTGAATAAATCTTCCACCGCTGTAACCTTAAGGTTAAGACAGTA 1667
|||||

Qy 61 CAAGATC 67
|||||
Db 1668 TAAATATC 1674

RESULT 4

AB091787	AB091787	5562 bp	DNA	linear	PRI 02-MAY-2003
LOCUS	Otolemur garnettii AMELX gene for amelogenin, partial cds.				
DEFINITION	Otolemur garnettii AMELX gene for amelogenin, partial cds.				
ACCESSION	AB091787				
VERSION	AB091787.1 GI:29126027				
KEYWORDS	.				
SOURCE	Otolemur garnettii (small-eared galago)				
ORGANISM	Otolemur garnettii				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Strepsirhini; Galagontidae; Otolemur.				
TITLE	1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N. From the Cover: The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)				
JOURNAL	2 (bases 1 to 5562) Iwase,M., Satta,Y. and Takahata,N. Direct Submission Submitted (19-Sep-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science; Shonan Kokusaiamura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwasemin@koryu.w01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)				
REFERENCE	Location/Qualifiers				
TITLE	1..5562				
DEFINITION	/organism="Otolemur garnettii"				
ACCESSION	/mol_type="genomic DNA"				
VERSION	/db_xref="taxon:30611"				
KEYWORDS	/sex="male"				
SOURCE	267..5562				
ORGANISM	/gene="AMELX"				
DEFINITION	John(267..321,1401..1412)				
AUTHORS	/gene="AMELX"				
TITLE	John(11413..1466,3385..3432,4699..4740,4832..4876,				
JOURNAL	5125..55562)				
REFERENCE	/gene="AMELX"				
TITLE	/codon_start=1				
DEFINITION	/product="amelogenin"				
ACCESSION	/protein_id="BAC66107.1"				
VERSION	/db_xref="GI:29126028"				
KEYWORDS	/translation="MGTWTLPACLGAAFPMLPPEHGHGYINFSYEKYSQAINID				
SOURCE	RVALVPLVKMYQSMLRPPYPSYSEEPGGVLLHHQLIVLSQQNPHTLQPHNHLIV				
ORGANISM	VPAQDPVVQPQPMHPVPGQSHMTPTONHQPMILPRPAQDPFQPCPQCPQPMOPMPOR				
DEFINITION	MQPMQPIQPQPFVHMQPLRPQPLRLPLRPWQPLRLPDLLEAWPAT"				
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					

FEATURES	source
gene	<p>1. .6931</p> <p>/organism="Sus scrofa"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9823"</p> <p>/sex="male"</p> <p>625. .6931</p> <p>/gene="AMELY"</p> <p>join(625. .680,2096. .2107)</p> <p>/gene="AMELY"</p> <p>join(2108. .2161,4806. .4853,6120. .6161,6218. .6262,6534. .>6931)</p> <p>/gene="AMELY"</p> <p>/codon_start=1</p> <p>/product="amelogenin"</p> <p>/protein_id="BAC6111.1"</p> <p>/db_xref="GI:29126037"</p> <p>/translation="MGTWIFPACILGASLAMPPEPHGPGYINFSYEDLYLEAICIDRTAFVLTPLMYQNMILPHYTSYXEBPMGMLHQLIPVVSQCTPQSHALQPHHIIIMPVPAQGPPIPOOPMPLPGHSMPTGHOHNPPLPAQGPPOPPVGPQHPQLPQSPMHP1QPLPQPLPPEPMSMGLPDLPLPMPATD"</p>
5'UTR	
CDS	
ORIGIN	
Query Match	55.5%; Score 37.2; DB 4; Length 6931;
Best Local Similarity	72.7%; Pred. No. 0.0069;
Matches	48; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
Oy	<p>1 CACAAGCTTGGAAATAAAGTCGCCCCACACAGTGTGAACCTTTAGGCTTTAAGACAGTA 60</p> <p>2187 CATTAAGCTTGGAAATAAATTCGCCCATAGATGTAACTTCAGGGCTTAAAAACAGA 2246</p>
Db	<p>61 CAAGAT 66</p> <p>2247 CCAGAT 2252</p>
RESULT 6	
LOCUS	AB091781 6465 bp DNA linear PRI 02-MAY-2003
DEFINITION	Pan troglodytes AMELX gene for amelogenin, partial cds.
ACCESSION	AB091781
VERSION	AB091781.1 GI:29126015
KEYWORDS	
SOURCE	
ORGANISM	<p>Pan troglodytes (chimpanzee)</p> <p>Pan troglodytes</p> <p>Eukaryota; Metazoa; Chordata; Cnidaria; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.</p>
REFERENCE	<p>1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N. From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)</p>
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	<p>12672962</p> <p>2 (bases 1 to 6931)</p>
REFERENCE	<p>Iwase,M., Satta,Y. and Takahata,N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokenai), Department of Biosystems Science; Shonan Kohnasimura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwase@minokoryu01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)</p>
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
PUBMED	<p>12672962</p> <p>2 (bases 1 to 6465)</p>
REFERENCE	<p>Iwase,M., Satta,Y. and Takahata,N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokenai), Department of Biosystems Science;</p>

Shonan kokusaijima, Hayama, Kanagawa 240-0193, Japan
 (E-mail: iwaseem@koryuw01.soken.ac.jp, Tel: 81-468-58-1571,
 Fax: 81-468-58-1544)

FEATURES

source

Location/Qualifiers

1. .6465

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/sex="male"

gene

927. .6465

/gene="AMELX"

5' UTR

Join(927. .382, 2283. .2294)

/gene="AMELX"

Join(2295. .2348, 4284. .4331, 5607. .5648, 5740. .5781, 6052. .56465)

/gene="AMELX"

/product="amelogenin"

/codon_start=1

/protein_id="BAC66101.1"

/db_xref="GI:29126016"

/translation="MGTWILFACLLGAAPFPPHGHGTYNSYENSHQATIND
 RTALVLTPLKWTQSIIRPTSTGIEPMGWLHQIIPVSOQPHPTLQPHHNPV
 PAQGPVLPQGPMPVPGHSMPTIOHQNPLPAPAQPPQPPVOPQPPQPPV
 HPWQPLPQPLPMPFMPQPLPPLPDLTLEAMPSTDKTK"

ORIGIN

Query Match 55.2%; Score 37; DB 9; Length 6465;
 Best Local Similarity 82.1%; Pred. No. 0.0082;
 Matches 55; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 1 CACAGCTTGAATAAAGTCTGCCACCACTTGAAGTTAGGTTAAGACAGTA 60
 |||||
 DB 2388 CACAGCTTGAATAAATCTGCC--ATACTGTGAATTAAGCTTAAACAGTA 2445

QY 61 CAAGATC 67
 |||||
 DB 2446 TGAGATC 2452

RESULT 7
 AB091791 7425 bp DNA linear MAM 02-MAY-2003
 LOCUS
 DEFINITION
 AB091791
 AB091791
 VERSION
 AB091791.1 GI:29126034
 KEYWORDS
 SOURCE
 ORGANISM
 Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 2 (bases 1 to 7425)
 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
 Direct Submission
 Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
 Advanced Studies (Soken), Kanagawa 240-0193, Japan
 (E-mail: iwaseem@koryuw01.soken.ac.jp, Tel: 81-468-58-1571,
 Fax: 81-468-58-1544)

FEATURES
 source
 1. .7425
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /sex="male"
 1491. .7425
 /gene="AMELX"

FEATURES

source

Location/Qualifiers

1. .7425

/organism="Sus scrofa"

/mol_type="genomic DNA"

/db_xref="taxon:9823"

/sex="male"

1491. .7425

/gene="AMELX"

gene

5' UTR
 CDS
 Join(1491. .1548, 3140. .3151)
 /gene="AMELX"
 Join(3152. .3205, 5316. .5363, 6619. .6660, 6717. .6761,
 7033. .57425)
 /gene="AMELX"
 /codon_start=1
 /product="amelogenin"
 /protein_id="BAC66110.1"
 /db_xref="GI:29126035"

ORIGIN

Query Match 54.6%; Score 36.6; DB 4; Length 7425;
 Best Local Similarity 71.6%; Pred. No. 0.012;
 Matches 48; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 CACAGCTTGAATAAAGTCTGCCACCACTTGAAGTTAGGTTAAGACAGTA 60
 |||||
 DB 3245 CATTACCTTGAATAAATCTACCCCGAATTGATACGTTAGGATGAACAGTA 3304

QY 61 CAAGATC 67
 |||||
 DB 3305 GAAGATC 3311

RESULT 8
 DB3729 2501 bp DNA linear PRI 10-JUL-2001
 LOCUS
 DEFINITION
 DB3729
 DB3729.1 GI:5263178
 VERSION
 DB3729.1 GI:5263178
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 2 (bases 1 to 2501)
 Sekiguchi, H., Minaguchi, K., Machida, Y. and Yakuhashi, M.
 PCR detection of the human amelogenin gene and its application to
 the diagnosis of amelo-genesis imperfecta
 Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 2 (bases 1 to 2501)
 Sekiguchi, H.
 Direct Submission
 Submitted (29-FEB-1996) Hiroshi Sekiguchi, Tokyo Dental College,
 Pediatric Dentistry, 1-2-2 Maeno, Minami-ku, Chiba 261,
 Japan (Tel: 043-270-3945, Fax: 043-279-2052)

FEATURES
 source
 1. .2501
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="Y"

gene
 1. .2501
 /gene="AMGY"
 1. .2501
 /note="AMGY"
 /note="amelogenin
 this nucleotide sequence shows an intron between exon2 and
 exon3 within human amelogenin gene on the Y chromosome"
 /number=2

ORIGIN

1. .7425

/organism="Sus scrofa"

/mol_type="genomic DNA"

/db_xref="taxon:9823"

/sex="male"

1491. .7425

/gene="AMELX"

Query Match 54.0%; Score 36.2; DB 9; Length 2501;
 Best Local Similarity 72.3%; Pred. No. 0.016;
 Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ACAGCTTGAATAAAGTCTGCCACCACTTGAAGTTAGGTTAAGACAGTAC 61

```

Db      39  ACAAAGTGAATTAAGATTGTCCTCCAGCTGCGAACTTAAGCTTTAAACAGCTT 98
QY      62  AGAT 66
Db      99  GAGCT 103

RESULT 9
AP000918/c 177654 bp DNA linear HTG 30-MAY-2000
LOCUS      Homo sapiens chromosome 18 clone RP11-737C10 map 18p11.3, WORKING
DEFINITION DRAFT SEQUENCE, 27 unordered pieces.
ACCESSION  AP000918
VERSION     AP000918.3 GI:8119059
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 177654)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
            Published Only in Database (1999)
            Homo sapiens 177,654 genomic DNA of 18p11.3
            Published Only in Database (1999)
            2 (bases 1 to 177654)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Matanabe,H. and Sakaki,Y.
            Direct Submission
            Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gs.c.riken.go.jp,
            URL:http://hgp.gs.c.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
            On May 31, 2000 this sequence version replaced gi:6997734.

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gs.c.riken.go.jp/
Contact: hattori@gs.c.riken.go.jp
----- Project Information
Center project name: HumDrafft2
Center clone name:
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 15691 bases at least Q40
Consensus quality: 15744 bases at least Q30
Consensus quality: 17231 bases at least Q20
Insert size: 175054; sum-of-coverage
Quality coverage: 4.38x in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of
27 contigs. The true order of the pieces is not known and the
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 46938 contig of 46938 bp in length
47039 69628 contig of 22590 bp in length
69729 77803 contig of 8075 bp in length
77904 86103 contig of 8200 bp in length
86204 94612 contig of 8409 bp in length
94713 103173 contig of 8461 bp in length
103274 110918 contig of 7645 bp in length
111019 117944 contig of 6926 bp in length
118045 124109 contig of 6065 bp in length
124210 129179 contig of 4970 bp in length
129280 136405 contig of 7126 bp in length
136506 142552 contig of 6047 bp in length
142653 146979 contig of 4327 bp in length
146980 151211 contig of 4132 bp in length
151212 155117 contig of 3806 bp in length
155118 155217 contig of 3662 bp in length
155218 158879 contig of 3662 bp in length
158880 158979 contig of 100 bp in length
158980 161806 contig of 2827 bp in length
161807 161906 contig of 100 bp in length
161907 164032 contig of 2126 bp in length
164033 164132 contig of 100 bp in length
164133 165745 contig of 1613 bp in length
165746 165845 contig of 100 bp in length
165846 167102 contig of 1257 bp in length
167103 167202 contig of 100 bp in length
167203 168212 contig of 1010 bp in length
168213 168312 contig of 1010 bp in length
168313 170352 contig of 1940 bp in length
170353 170352 contig of 100 bp in length
170353 171734 contig of 1382 bp in length
171735 171834 contig of 100 bp in length
171835 173810 contig of 1976 bp in length
173811 173910 contig of 100 bp in length

```

```

142653 146979 contig of 4327 bp in length
147080 151211 contig of 4132 bp in length
151312 155117 contig of 3806 bp in length
155218 158879 contig of 3662 bp in length
158980 161806 contig of 2827 bp in length
161907 164032 contig of 2126 bp in length
164133 165745 contig of 1613 bp in length
165846 167102 contig of 1010 bp in length
167203 168212 contig of 1010 bp in length
168313 170352 contig of 1940 bp in length
170353 171734 contig of 1382 bp in length
171835 173810 contig of 1976 bp in length
173911 175003 contig of 1093 bp in length
175104 176492 contig of 1389 bp in length
176593 177654 contig of 1062 bp in length

Sequence updated (06-Jan-2000)
Sequence updated (26-May-2000).
NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 46938 contig of 46938 bp in length
46939 47038 contig of 100 bp
47039 69628 contig of 22590 bp in length
69629 69728 contig of 100 bp
69729 77803 contig of 8075 bp in length
77804 77903 contig of 100 bp
77904 86103 contig of 8200 bp in length
86104 86203 contig of 100 bp
86204 94612 contig of 8409 bp in length
94613 94712 contig of 100 bp
94713 103173 contig of 8461 bp in length
103174 103273 contig of 100 bp
103274 110918 contig of 7645 bp in length
110919 111018 contig of 100 bp
111019 117944 contig of 6926 bp in length
117945 118044 contig of 100 bp
118045 124109 contig of 6065 bp in length
124110 124209 contig of 100 bp
124210 129179 contig of 4970 bp in length
129180 129279 contig of 100 bp
129280 136405 contig of 7126 bp in length
136406 136505 contig of 100 bp
136506 142552 contig of 6047 bp in length
142553 142652 contig of 100 bp
142653 146979 contig of 4327 bp in length
146980 147079 contig of 100 bp
147080 151211 contig of 4132 bp in length
151212 151311 contig of 100 bp
151312 155117 contig of 3806 bp in length
155118 155217 contig of 100 bp
155218 158879 contig of 3662 bp in length
158880 158979 contig of 100 bp
158980 161806 contig of 2827 bp in length
161807 161906 contig of 100 bp
161907 164032 contig of 2126 bp in length
164033 164132 contig of 100 bp
164133 165745 contig of 1613 bp in length
165746 165845 contig of 100 bp
165846 167102 contig of 1257 bp in length
167103 167202 contig of 100 bp
167203 168212 contig of 1010 bp in length
168213 168312 contig of 1010 bp in length
168313 170352 contig of 1940 bp in length
170353 170352 contig of 100 bp
170353 171734 contig of 1382 bp in length
171735 171834 contig of 100 bp
171835 173810 contig of 1976 bp in length
173811 173910 contig of 100 bp

```

FEATURES

* 173911 175003: contig of 1093 bp in length

* 175004 175103: gap of 100 bp

* 175104 176492: contig of 1389 bp in length

* 176493 176592: gap of 100 bp

* 176593 177654: contig of 1062 bp in length.

source

1. 177654

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="18"

/map="18p11.3"

/clone="RP11-737C10"

1. 46938

/note="assembly_fragment"

47039. 69628

/note="assembly_fragment"

69729. 77803

/note="assembly_fragment"

77904. 86103

/note="assembly_fragment"

86204. 94612

/note="assembly_fragment clone_end:17 vector_side:right"

94713. 103173

/note="assembly_fragment"

103274. 110918

/note="assembly_fragment"

111019. 117944

/note="assembly_fragment"

118045. 124109

/note="assembly_fragment"

124210. 129179

/note="assembly_fragment"

129280. 136405

/note="assembly_fragment"

136506. 142552

/note="assembly_fragment"

142653. 146979

/note="assembly_fragment"

147080. 151211

/note="assembly_fragment"

151312. 155117

/note="assembly_fragment"

155218. 158879

/note="assembly_fragment"

158980. 161806

/note="assembly_fragment"

161907. 164032

/note="assembly_fragment"

164133. 165745

/note="assembly_fragment"

165846. 167102

/note="assembly_fragment"

167203. 168212

/note="assembly_fragment"

168313. 170252

/note="assembly_fragment clone_end:SP6 vector_side:right"

170353. 171734

/note="assembly_fragment"

171835. 173810

/note="assembly_fragment"

173911. 175003

/note="assembly_fragment"

175104. 176492

Query Match 54.0%; Score 36.2; DB 2; Length 177654;

Best Local Similarity 72.3%; Pred. No. 0.017;

Matches 4; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ACAAGCTGGAATTAAGTCTGCCCAACAGTTGTAAGTTTGGGTTTAAGACAGTAC 61

DB 63974 ACAAACTGGAATTAAGTCTGCCCAACAGTTGTAAGTTTGGGTTTAAACAGTCTT 63915

QY 62 AGAT 66

DB 63914 GAGCT 63910

RESULT 10

AC013412/c

LOCUS

DEFINITION

AC013412

AC013412

VERSION

AC013412.3 GI:8570260

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 200214)

1 (bases 1 to 200214)

Sullivan, J.B. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

MEDLINE

9847074

2 (bases 1 to 200214)

REFERENCE

Bemis, G., Kallio, J., Stoneking, T. and Parker, C.

The sequence of Homo sapiens BAC clone RP11-507A3

Unpublished

3 (bases 1 to 200214)

Waterston, R.H.

Direct Submission

Submitted (09-NOV-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 200214)

Waterston, R.H.

Direct Submission

Submitted (17-JUN-2000) Genome Sequencing Center, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

MO 63108, USA

5 (bases 1 to 200214)

Waterston, R.

Direct Submission

Submitted (30-SEP-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Jun 17, 2000 this sequence version replaced gi:7631033.

REFERENCE

Waterston, R.

Direct Submission

Submitted (30-SEP-2000) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Jun 17, 2000 this sequence version replaced gi:7631033.

COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.edu

Summary Statistics

Center project name: H_NH0507A03

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skalecky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:

The RPL11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frenken, E., Tateo, M., Catanesi, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RPL1-11SH13, 200 bp overlap. Actual start of this clone is at base position 1 of RPL1-507A3. Actual end is at base position 20020 of RPL1-507A3.

Location/Qualifiers

```
FEATURES
source
1..200214
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RPL1-507A3"
/clone_1lb="RPL1-11"
113..313
/rpt_family="L2"
598..666
/rpt_family="L1"
668..4324
/rpt_family="L1"
4325..4743
/rpt_family="MALR"
4744..6332
/rpt_family="L1"
6365..6645
/rpt_family="L1"
6763..7351
/rpt_family="L1"
7378..7691
/rpt_family="L2"
7759..8261
/rpt_family="L2"
8595..8698
/rpt_family="MER1_type"
9291..9536
/rpt_family="MIR"
9872..10006
/rpt_family="MER1_type"
10795..11093
/rpt_family="Alu"
12150..12586
/rpt_family="L1"
12915..13065
/rpt_family="L1"
13795..14551
/rpt_family="L1"
15020..15724
/rpt_family="L1"
15725..16025
/rpt_family="Alu"
16026..16736
/rpt_family="L1"
16835..17333
/rpt_family="L1"
17536..18235
/rpt_family="L1"
19412..19593
/rpt_family="ERV1"
19911..20152
/rpt_family="MALR"
20339..20440
/rpt_family="MIR"
20441..20971
/rpt_family="MALR"
repeat_region
```

```
/rpt_family="ERV1"
20972..21394
/rpt_family="MALR"
23267..23436
/note="similar to EST AW662994 (NID:G7455535) hh83h05.y1"
misc_feature
23276..23436
/note="similar to EST AI439133 (NID:G4302189) t187e02.x1"
misc_feature
23278..23436
/note="similar to EST AI089900 (NID:G3428959) qa16a04.x1"
misc_feature
23283..23436
/note="similar to EST AW090465 (NID:G6047809) xc83e02.x1"
misc_feature
23484..23745
/note="similar to EST AW090465 (NID:G6047809) xc83e02.x1"
misc_feature
23484..23773
/note="similar to EST AL039198 (NID:G5408274) "
misc_feature
23768..23988
/note="similar to EST AI570004 (NID:G4533378) tr90h07.x1"
repeat_region
23768..23988
/rpt_family="ERV1"
24059..24874
/rpt_family="ERV1"
25041..25686
/rpt_family="ERV1"
25872..26379
/rpt_family="ERV1"
26379..26379
/rpt_family="ERV1"
32050..32190
/rpt_family="MIR"
35963..36087
/rpt_family="L1"
36190..36324
/rpt_family="L2"
39009..39268
/rpt_family="ERV1"
39269..39542
/rpt_family="Alu"
39558..39702
/rpt_family="L1"
39722..39880
/rpt_family="L1"
39913..40227
/rpt_family="L1"
40421..40560
/rpt_family="L1"
40561..40633
/rpt_family="Alu"
40634..40839
/rpt_family="Alu"
40857..41158
/rpt_family="L1"
41312..41413
/rpt_family="CR1"
41435..41483
/rpt_family="L2"
42963..44805
/rpt_family="L1"
44804..45325
/rpt_family="L1"
45384..45445
/rpt_family="L1"
45446..45757
/rpt_family="Alu"
45758..45797
/rpt_family="L1"
repeat_region
```

Query Match 54.0%; Score 36.2; DB 9; Length 200214;

Best Local Similarity 72.3%; Pred. No. 0.017;

Matches 47; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
QY 2 ACAAGCTGGAATTAAGCTGCCCCACAGTGTGTAAGCTTTAAGACAGTAC 61
DB 177037 ACAACGTGGAATTAAGCTGCCCCACAGTGTGTAAGCTTTAAGACAGT 176978
QY 62 AAGAT 66
DB 176977 GAGCT 176973
```

RESULT 11
D83730 1935 bp DNA 1linear PRI 10-JUL-2001
LOCUS D83730
DEFINITION Homo sapiens AMGX gene for amelogenin, intron 2.
ACCESSION D83730.1 GI:5263179
VERSION D83730.1
KEYWORDS amelogenin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1
TITLE Sekiguchi, H., Minaguchi, K., Machida, Y. and Yakushiji, M.
JOURNAL PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta
MEDLINE Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)
PUBMED 99234629
10218009
REFERENCE 2 (bases 1 to 1935)
AUTHORS Sekiguchi, H.
TITLE Direct Submission
JOURNAL Submitted (29-FEB-1996) Hiroshi Sekiguchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Minama-ku, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)
Location/Qualifiers
1. .1935
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"

FEATURES
source
1. .1935
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"

gene
1. .1935
/gene="AMGX"
1. .1935
/gene="AMGX"
/note="amelogenin
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"
/number=2

ORIGIN
Query Match 52.8%; Score 35.4; DB 9; Length 1935;
Best Local Similarity 80.6%; Pred. No. 0.032;
Matches 54; Conservative 0; Mismatches 11; Indels 2; Gaps 1;
QY 1 CACAAAGCTTGAATAAAGTCTGCCACCAAGTTGTAACCTTAGGTTTAAAGACAGTA 60
DB 40 CACAAAGCTTGAATAAAGTCTGCC-ATAGTTGTGAATAATGAGTTTAAACAGTA 97
QY 61 CAAGATC 67
DB 98 TGAGATC 104

RESULT 12
AB091794 5591 bp DNA 1linear MAM 02-MAY-2003
LOCUS AB091794
DEFINITION Equus caballus AMELX gene for amelogenin, partial cds.
ACCESSION AB091794
VERSION AB091794.1 GI:29126040
KEYWORDS
SOURCE Equus caballus (horse)
ORGANISM Equus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

REFERENCE
AUTHORS 1
TITLE Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
JOURNAL From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
PUBMED 12672962

REFERENCE 2 (bases 1 to 5591)
AUTHORS Iwase, M., Satta, Y. and Takahata, N.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Soken), Department of Biosystems Science; Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES
source
1. .5591
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/sex="male"
130. .5591
/gene="AMELY"
/gene="AMELY"
/join(130. .180, 1493. .1494)
/join(1495. .1548, 3470. .3517, 4907. .4915, 5235. .5591)
/gene="AMELY"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66113.1"
/db_xref="GI:29126041"
/translation="MGTWILFACLVGTAIAMPLRPHGHPAYINSEVLTYPYSGYE
PMSGHQIIPVLTQONPSNALQPHHHPVPAQGVSVSOQPAIPLEGHSMIPSQ
HHQPLRPVPPVQPPQPHQPIQPPRPMHPICPLPQAPLPPIPLAASGPHAS"

ORIGIN
Query Match 52.2%; Score 35; DB 4; Length 5591;
Best Local Similarity 70.1%; Pred. No. 0.046;
Matches 47; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 1 CACAAAGCTTGAATAAAGTCTGCCACCAAGTTGTAACCTTAGGTTTAAAGACAGTA 60
DB 1574 CACAAAGCTTGAATGCATAAATGCGCTCATTTGTGTAAATTTGGGGTTTAATCAATA 1633
QY 61 CAAGATC 67
DB 1634 CAAGATC 1640

RESULT 13
AB091783 6442 bp DNA 1linear PRI 02-MAY-2003
LOCUS AB091783
DEFINITION Saimiri sciureus AMELX gene for amelogenin, partial cds.
ACCESSION AB091783
VERSION AB091783.1 GI:29126019
KEYWORDS
SOURCE Saimiri sciureus (common squirrel monkey)
ORGANISM Saimiri sciureus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.

REFERENCE
AUTHORS 1
TITLE Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
JOURNAL From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
PUBMED 12672962
12672962
REFERENCE 2 (bases 1 to 6442)
AUTHORS Iwase, M., Satta, Y. and Takahata, N.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Soken), Department of Biosystems Science; Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES
source
1. .6442
/organism="Saimiri sciureus"
/mol_type="genomic DNA"

```

/db_xref="taxon:9521"
/sex="male"
936..6442
/gene="AMELX"
/join(936..991,2298..2309)
/gene="AMELX"
/join(2310..2363,4289..4336,5592..5633,5725..5769,
6041..56442)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66103.1"
/db_xref="GI:29126020"
/translation="MGTWILFACILGAAPAMP.LPHPGHGTYNFSYENSHQALNID
KRLVUTPLKMYQSMIRPVPYSGYRPMGMLHQIIPVISOQHPLTHLQSHHLPV
VPAQOPVPEQPMVPVPGQSMPTQHHQPNPLPVAQOPFQVQPPVQPPHQPQPP
VHMQPLPQPPLPMPFPMQPLPPLPDLLEAWPAT"

ORIGIN

Query Match          51.6%; Score 34.8; DB 9; Length 6442;
Best Local Similarity 80.3%; Pred. No. 0.054;
Matches 53; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 CACAACTTGAATTAAGCTGCCCCACCACTTGCTTAAGGTTTAAGACAGTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2401 CATAAGATTGCAATTAAGATTCTGCCCA-TGGTTGTTGAAGTTTAAGACAGTA 2459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 CAAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2460 TGAGAT 2465

RESULT 14
AB091782 7163 bp DNA linear PRI 02-MAY-2003
LOCUS Pan troglodytes AMELY gene for amelogenin, partial cds.
DEFINITION AB091782
ACCESSION AB091782
VERSION AB091782.1 GI:29126017
KEYWORDS
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS Iwase,M., Saito,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
TITLE boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569
PUBMED 12672962
2 (bases 1 to 7163)
REFERENCE Iwase,M., Saito,Y. and Takahata,N.
AUTHORS Direct Submision
TITLE Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
JOURNAL Advanced Studies(Sokendai), Department of Biosystems Science;
Shonan Kikusaijima, Hayama, Kanagawa 240-0193, Japan
(E-mail:iwase@koryu.w01.soken.ac.jp, Tel:81-468-58-1571,
Fax:81-468-58-1544)
Location/Qualifiers
1. 7163
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/sex="male"
1186..7163
/gene="AMELY"
/join(1186..1241,2593..2604)
/gene="AMELY"
/join(2605..2658,5151..5198,6297..6338,6430..6474,
6744..7163)
/gene="AMELY"
/codon_start=1
/product="amelogenin"

FEATURES
source
gene
5'UTR
CDS

```

```

/protein_id="BAC66102.1"
/db_xref="GI:29126018"
/translation="MGTWILFACILGAAPAMP.LPHPGHGTYNFSYENSHQALNID
KRLVUTPLKMYQSMIRPVPYSGYRPMGMLHQIIPVISOQHPLTHLQSHHLPV
VPAQOPVPEQPMVPVPGQSMPTQHHQPNPLPVAQOPFQVQPPVQPPHQPQPP
VHMQPLPQPPLPMPFPMQPLPPLPDLLEAWPATDKRKE"

ORIGIN

Query Match          51.6%; Score 34.6; DB 9; Length 7163;
Best Local Similarity 70.8%; Pred. No. 0.064;
Matches 46; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 ACAAGCTGGAATTAAGCTGCCACAGTGTGAAGTTTAAGGTTTAAGACAGTAC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3732 ACAAGCTGGAATTAAGATTGTCCCCAGCTGGAAACTTTAAGTTTAAGACAGTTT 3673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 AAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3672 GAGCT 3668

RESULT 15
BS000568/c 38765 bp DNA linear PRI 17-MAR-2004
LOCUS Pan troglodytes chromosome Y clone:PTFY-001K12, complete sequences.
DEFINITION BS000568
ACCESSION BS000568
VERSION BS000568.1 GI:45504168
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
AUTHORS The Chimpanzee Chromosome Y Sequencing Consortium.
TITLE DNA sequence of chimpanzee chromosome Y and its evolutionary
JOURNAL implications
AUTHORS 2 (bases 1 to 38765)
Hattori,M., Toyoda,A., Noguchi,H., Taylor,T.D., Kuroki,Y.,
Fujiyama,A. and Sakaki,Y.
TITLE Direct Submision
JOURNAL Submitted (15-MAR-2004) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Location/Qualifiers
1. 38765
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="Y"
/clone="PTFY-001K12"

FEATURES
source
ORIGIN

Query Match          51.6%; Score 34.6; DB 9; Length 38765;
Best Local Similarity 70.8%; Pred. No. 0.064;
Matches 46; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 ACAAGCTGGAATTAAGCTGCCACAGTGTGAAGTTTAAGGTTTAAGACAGTAC 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3732 ACAAGCTGGAATTAAGATTGTCCCCAGCTGGAAACTTTAAGTTTAAGACAGTTT 3673
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 AAGAT 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3672 GAGCT 3668

Search completed: February 8, 2005, 16:36:59
Job time : 377.57 secs

```

This Page Blank (uspto)

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 50.4%; Score 33.8; DB 12; Length 259202;

Best Local Similarity 79.1%; Pred. No. 0.045; Indels 2; Gaps 1;

Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

Db 69966 CACAACTTGACATTAATAATCTGCTC-ATAGTTGTAATTAAGGCTTTAAACAGTA 70023

QY 61 CAGATC 67

Db 70024 TGAGATC 70030

RESULT 2

ABV98228/c

ID ABV98228 standard; cDNA; 559 BP.

XX ABV98228;

XX 14-JAN-2003 (first entry)

DE Human pancreatic cancer expressed cDNA SEQ ID NO 3636.

KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;

KW cytostatic; tumour; gene; ss.

XX Homo sapiens.

XX MO200260317-A2.

XX 08-ANG-2002.

XX 30-JAN-2002; 2002MO-US002781.

XX 30-JAN-2001; 2001US-0265305P.

XX 31-JAN-2001; 2001US-0265682P.

XX 09-FEB-2001; 2001US-0267588P.

XX 21-MAR-2001; 2001US-0278651P.

XX 28-APR-2001; 2001US-0287112P.

XX 16-MAY-2001; 2001US-0291631P.

XX 12-JUL-2001; 2001US-0305484P.

XX 20-AUG-2001; 2001US-0313999P.

XX 27-NOV-2001; 2001US-0333626P.

XX (CORI-) CORIXA CORP.

XX Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;

XX WPI; 2002-627435/67.

XX New isolated polynucleotide and pancreatic tumor polypeptides, useful for

XX diagnosing, preventing and/or treating cancer, particularly pancreatic

XX cancer.

XX Claim 1, SEQ ID NO 3636; 300bp + Sequence Listing; English.

CC cells expressing the polypeptide are useful in treating pancreatic cancer
CC and stimulating an immune response. The polynucleotides can be used as
CC probes or primers for nucleic acid hybridization, in the design and
CC preparation of ribozyme molecules for inhibiting expression of the tumour
CC polypeptide and proteins in the tumour cells, in vaccines and for gene
CC therapy. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 559 BP; 193 A; 84 C; 111 G; 171 T; 0 U; 0 Other;

Query Match 39.7%; Score 26.6; DB 6; Length 559;

Best Local Similarity 63.1%; Pred. No. 5.7;

Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Db 293 CAAATATGATTAATAATGTTGTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCA 234

QY 63 AGATC 67

Db 233 TGATC 229

RESULT 3

ABO60609

ID ABO60609 standard; cDNA; 614 BP.

XX ABO60609;

XX 02-AUG-2002 (first entry)

DE Human colon cancer related nucleotide sequence SEQ ID NO:4304.

KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;

KW genetic analysis; diagnostic; antisense therapy; gene; ss.

XX Homo sapiens.

XX MO200229086-A2.

XX 11-APR-2002.

XX 02-OCT-2001; 2001MO-US030732.

XX 02-OCT-2000; 2000US-0237271P.

XX (FARB) BAYER CORP.

XX Burgess C, Astle JH, Carroll E, Catino TV, Dwivedi P, Molino GA;

XX Thagilangam A, Lewis ME;

XX WPI; 2002-426115/45.

XX New isolated nucleic acid that is differentially expressed in cancer

XX tissues useful for determining the presence of colon cancer in a cell or

XX tissue type, and in antisense therapy.

XX Claim 1, Fig 1; 796bp; English.

XX ABO56306 to ABO60787 represent isolated nucleic acids (I) differentially

XX expressed in cancer tissues. ABO7893 to ABO7904 represent proteins

XX encoded by the ABO60776 to ABO60787 nucleic acid sequences. (I) can be

XX used in antisense therapy. An antibody immunoreactive with a polypeptide

XX encoded by (I) is useful for detecting cancer in a patient sample, and

XX for detecting the presence or absence of a polynucleotide encoded by a

XX nucleic acid which hybridizes to (I) in a cell. A probe/primer derived

XX from (I) can be used for determining the presence of a nucleic acid which

XX hybridizes to (I), and for determining the phenotype of cells in a sample

XX of cells from a patient. (I) is useful for determining the presence of

XX colon cancer in a cell or tissue type, for determining the presence of

XX state of other type of cancer, in antisense therapy, to generate

XX microarrays on a solid surface, to identify a chromosome on which the

CC corresponding gene residues, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
XX

SQ Sequence 614 BP; 186 A; 126 C; 81 G; 216 T; 0 U; 5 Other;

Query Match 39.7%; Score 26.6; DB 6; Length 614;
Best Local Similarity 63.1%; Pred. No. 5.8; 24; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATTAATAAGTCTGCCACAGTTGTTACTTTAGGCTTTAAGACACTACA 62
DB 261 CAATATGATTAATTAATGTTTGCGCCCGAGTTGTTATCTCTAAGATTAAGACTATCA 320

QY 63 AGATC 67
DB 321 TGATC 325

RESULT 4

ADP12803/c
ID ADP12803 standard; DNA; 2176 BP.

AC ADP12803;

DT 12-AUG-2004 (first entry)

DE Reference mRNA sequence #17.

XX transplant rejection; immune system; rheumatoid arthritis; lupus;

KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.

XX Homo sapiens.

PN WO2004042346-A2.

PD 21-MAY-2004.

PF 24-APR-2003; 2003WO-US012946.

PR 24-APR-2002; 2002US-00131831.

PR 20-DEC-2002; 2002US-00325899.

PA (EXPR-) EXPRESSION DIAGNOSTICS INC.

PI Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;

PI Rosenberg S;

DR MPI; 2004-400724/37.

XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
PT rejection, in an individual, comprises detecting the expression level of
PT the genes.

PS Claim 65; SEQ ID NO 2812; 1762bp; English.

CC The present invention relates to diagnosing or monitoring transplant
CC rejection, e.g. cardiac or kidney transplant rejection, in an individual
CC comprising detecting the expression level of one or more genes. The
CC methods, system and kits are useful in diagnosing or monitoring
CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
CC islet, lung, bone marrow or stem cell transplant rejection,
CC xenotransplant rejection or mechanical organ replacement rejection, in an
CC individual. The method is also useful in assessing the immune status of
CC an individual. The methods are also useful in diagnosing and monitoring
CC diseases that involve the immune system, e.g. rheumatoid arthritis,
CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
CC viral, bacterial or fungal infection. The present sequence represents a
CC reference mRNA sequence of the invention which show altered expression in
CC renal transplantation and expression.

XX Sequence 2176 BP; 770 A; 309 C; 425 G; 672 T; 0 U; 0 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 2176;
Best Local Similarity 63.1%; Pred. No. 7.9; 24; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATTAATAAGTCTGCCACAGTTGTTACTTTAGGCTTTAAGACACTACA 62
DB 1471 CAATATGATTAATTAATGTTTGCGCCCGAGTTGTTATCTCTAAGATTAAGACTATCA 1412

QY 63 AGATC 67
DB 1411 TGATC 1407

RESULT 5

ADQ24473/c
ID ADQ24473 standard; DNA; 2516 BP.

AC ADQ24473;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 7293.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;

KW ds.

XX Homo sapiens.

PN WO2004048938-A2.

PD 10-JUN-2004.

PF 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

PA (PROT-) PROTEIN DESIGN LABS INC.

PI Aziz N, Ginsburg WM, Zlotnik A;

DR MPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

PS Example 2; SEQ ID NO 7293; 210bp; English.

CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 2516 BP; 862 A; 376 C; 484 G; 779 T; 0 U; 15 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 2516;
Best Local Similarity 63.1%; Pred. No. 8.1; 24; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATTAATAAGTCTGCCACAGTTGTTACTTTAGGCTTTAAGACACTACA 62
DB 1665 CAATATGATTAATTAATGTTTGCGCCCGAGTTGTTATCTCTAAGATTAAGACTATCA 1606

OY 63 AGATC 67
DB 1605 TGATC 1601

RESULT 6

ADQ22895/c
ID ADQ22895 standard; DNA; 2516 BP.

XX
AC ADQ22895;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 5715.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM db.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX Example 2; SEQ ID NO 5715; 210pp; English.

CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 2516 BP; 862 A; 376 C; 484 G; 779 T; 0 U; 15 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 2516;

Best Local Similarity 63.1%; Pred. No. 8.1; Mismatches 24; Indels 0; Gaps 0;

OY 3 CAAGCTTGAATAAAGTCTGCCCCACAGTTGGTAACCTTAGGGTTTAAGACAGTACA 62
DB 1665 CAATATATGATTAATGTTGTGCCCCCAGTTTATCTTAAGATAGATATCA 1606

OY 63 AGATC 67

DB 1605 TGATC 1601

RESULT 7

AD161706/c

ID AD161706 standard; cDNA; 9262 BP.

XX AD161706;

XX 22-APR-2004 (first entry)

XX Human cDNA downregulated in Alzheimer's disease, INCYTE 24655.92.

XX Human; ss; Alzheimer's disease; differential display; neuroprotective;
XX brain disorder.

XX Homo sapiens.

XX US6682888-B1.

XX 27-JAN-2004.

XX 05-MAY-2000; 2000US-00566921.

XX 05-MAY-2000; 2000US-00566921.

XX (INCY-) INCYTE CORP.

XX Loring JF, Tingley DW, Edwards CM;

XX WPI; 2004-118572/12.

XX New composition comprising cDNAs that are differentially expressed in
PT brain disorders, useful for diagnosing or treating Alzheimer's disease.
XX Claim 1; SEQ ID NO 74; 223pp; English.

CC The invention relates to a new composition comprising AD161633-
CC AD161706 and their complements that are cDNAs differentially expressed in
CC brain disorders. Also included are a high throughput method for detecting
CC differential expression of one or more cDNAs in a sample containing
CC nucleic acids and a high throughput method for screening a library of
CC molecules or compounds to identify a ligand that specifically binds a
CC cDNA. The expression of the each of the cDNAs is downregulated at least
CC two-fold in the brain of the subjects with Alzheimer's disease (AD161727-
CC AD161727) or upregulated at least two fold in Alzheimer's disease
CC (AD161728-AD161770). The composition is useful for diagnosing or treating
CC Alzheimer's disease. The present sequence is a cDNA downregulated at
CC least two-fold in the brain of the subjects with Alzheimer's disease.

XX Sequence 9262 BP; 2795 A; 1687 C; 1820 G; 2951 T; 0 U; 9 Other;

Query Match 39.7%; Score 26.6; DB 12; Length 9262;

Best Local Similarity 63.1%; Pred. No. 11; Mismatches 24; Indels 0; Gaps 0;

OY 3 CAAGCTTGAATAAAGTCTGCCCCACAGTTGGTAACCTTAGGGTTTAAGACAGTACA 62
DB 5902 CAATATATGATTAATGTTGTGCCCCCAGTTTATCTTAAGATAGATATCA 5843

OY 63 AGATC 67

DB 5842 TGATC 5838

RESULT 8

ADJ56289/c

XX ADJ56289 standard; cDNA; 9277 BP.

XX ADJ56289;

XX 06-MAY-2004 (first entry)

XX Human cDNA differentially expressed in MYCN activated cells SeqID 95.
XX human; differential expression; transactivator; proto-oncogene;
XX neuroblastoma; small cell lung cancer; cytostatic; gene therapy; ss;
XX MYCN activated cell.


```

XX OS Homo sapiens.
XX PN US2003119009-A1.
XX PD 26-JUN-2003.
XX PF 25-FEB-2002; 2002US-00084817.
XX PR 23-FEB-2001; 2001US-0270784P.
XX PA (STUA/) STUART S G.
XX PA (NUCHT) NUCHTERN J G.
XX PA (PLON/) PLON S E.
XX PA (SHOH/) SHOHEI J M.
XX PI Stuart SG, Nuchtern JG, Plon SE, Shohe J M;
XX DR MPI; 2003-635698/60.
XX PT New genes regulated by MYCN activation, useful in gene therapy,
XX PT particularly for treating a subject with e.g. neuroblastoma or other
XX PT cancers, or for diagnosing, staging or monitoring the treatment of the
XX PT cancer.
XX PS Claim 1; SEQ ID NO 95; 27bp; English.
XX CC This invention relates to novel isolated cDNAs that are differentially
XX CC expressed in MYCN activated cells. Specifically, it refers to
XX CC polynucleotide sequences that exhibit differential expression patterns in
XX CC cells activated by the transactivator MYCN, where MYCN is a proto-
XX CC oncogene that is amplified in neuroblastoma cells and is common in small
XX CC cell lung cancers. The present invention describes these cDNA molecules
XX CC as useful for in hybridization assays to detect expression of nucleic
XX CC acids (or complementary nucleic acids) in a present in a given sample, as
XX CC well as for screening assays by identifying molecules or compounds that
XX CC specifically bind the cDNA as a ligand and modulate function or activity.
XX CC Accordingly, these compositions exhibit cytostatic activity and can also
XX CC be used for gene therapy purposes. This polynucleotide sequence is a cDNA
XX CC that is differentially expressed in MYCN activated cells, given in an
XX CC exemplification of the invention. NOTE: This sequence does not appear in
XX CC the printed specification but has been obtained in electronic format from
XX CC the US Patent Office at:
XX CC ftp.segdata.neptco.gov/sequence.html?docid=20030119009.
XX SQ Sequence 9277 BP; 2805 A; 1687 C; 1820 G; 2955 T; 0 U; 10 Other;
XX
XX Query Match 39.7%; Score 26.6; DB 10; Length 9277;
XX Best Local Similarity 63.1%; Pred. No. 11;
XX Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 3 CAAGCTTGAAATAAAGTCTGCCACAGTTGTAAGGTTTAAAGACAGTACA 62
DB 5902 CATATATGATTAATTAATGTTGTGCCCCGAGTTGTTATCTTAAGATTAAGATTAACA 5843
QY 63 AGATC 67
DB 5842 TGATC 5838
XX
XX RESULT 9
XX ADC86176
XX ID ADC86176 standard; DNA; 47804 BP.
XX AC ADC86176;
XX XX
XX DT 01-JAN-2004 (first entry)
XX XX
XX DE Human GPCR gene SEQ ID NO:629.
XX XX
XX KW de; gene; human; GPCR;
XX KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX XX

```

```

OS OS Homo sapiens.
XX XX EPI270724-A2.
XX XX 02-JAN-2003.
XX PD 18-JUN-2002; 2002EP-00013517.
XX PF 18-JUN-2001; 2001JP-00246789.
XX PR 18-JUN-2001; 2001JP-00246789.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATION.
XX PI Suwa M, Aaei K, Akiyama Y, Aburatani H;
XX DR MPI; 2003-315783/31.
XX DR P-PSDB; ADC86177.
XX PT New polynucleotide, useful for preparing a composition for treating a
XX PT patient in need of increased or suppressed activity or expression of the
XX PT guanosine triphosphate-binding protein coupled receptor.
XX PS Claim 1; SEQ ID NO 629; 28bp; English.
XX CC The invention relates to a novel polynucleotide encoding a guanosine
XX CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX CC the invention may have a use in gene therapy. The polynucleotide and
XX CC polypeptide are useful for preparing a composition for treating a patient
XX CC in need of increased or suppressed activity or expression of the
XX CC guanosine triphosphate-binding protein coupled receptor. The
XX CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
XX CC invention.
XX SQ Sequence 47804 BP; 13080 A; 10916 C; 11116 G; 12192 T; 0 U; 500 Other;
XX
XX Query Match 39.1%; Score 26.2; DB 10; Length 47804;
XX Best Local Similarity 63.5%; Pred. No. 23;
XX Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 4 AAGCTTGAAATAAAGTCTGCCACAGTTGTAAGGTTTAAAGACAGTACA 63
DB 38451 AAACCTTAAGTAATTAAGAGCCCCCCCCATGACAGTGTAACTTATGATTAATGATTAAGACAT 38510
QY 64 GAT 66
DB 38511 GAT 38513
XX
XX RESULT 10
XX ACH13083/C
XX ID ACH13083 standard; cDNA; 455 BP.
XX AC ACH13083;
XX XX
XX DT 13-OCT-2003 (first entry)
XX XX
XX DE Human adult brain cDNA #295.
XX XX
XX KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX KW genome mapping; biodiversity; genetic disorder.
XX XX
XX OS Homo sapiens.
XX PN US2003073623-A1.
XX PD 17-APR-2003.
XX PF 30-JUL-2001; 2001US-00918995.
XX PR 30-JUL-2001; 2001US-00918995.
XX PA (DRMA/) DRMANAC R T.
XX PA (LABA/) LABAT I.
XX XX

```



```
FT FT /tag= 0
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(3882,TT)
FT FT **tag= P
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(3988,T)
FT FT **tag= G
FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace(4452,A)
FT FT **tag= T
FT FT /standard_name= "Single nucleotide polymorphism"
XX XX US2003190617-A1.
XX XX 09-OCT-2003.
XX XX 06-MAR-2002; 2002US-00091281.
XX XX 06-MAR-2002; 2002US-00091281.
XX XX (SIEE/) SI E.
XX XX (RAYM/) RAYMOND V.
XX XX (MORI/) MORISSETTE J.
XX XX Raymond V, Morissette J, Si E;
XX XX WPI; 2003-864168/80.
XX XX
XX XX New nucleic acid sequences of the optineurin gene are useful to detect
XX XX polymorphisms particularly single nucleotide polymorphisms in the
XX XX optineurin promoter to diagnose, prognose and treat glaucoma and related
XX XX disorders.
XX XX Disclosure; SEQ ID NO 2; 159pp; English.
XX XX
XX XX The invention relates to an isolated nucleic acid (NI) comprising at
XX XX least 20 but not more than 1500 consecutive nucleotides of the optineurin
XX XX promoter appearing as ADEI3890. Also included are the optineurin promoter
XX XX operably linked to a heterologous nucleic acid, a nucleic acid capable of
XX XX detecting a single nucleotide polymorphism (SNP) in the optineurin
XX XX promoter, a host cell comprising the promoter operably linked to a
XX XX heterologous sequence, diagnosing or prognosing glaucoma in a sample
XX XX obtained from a cell or bodily fluid (comprising detecting a polymorphism
XX XX in a promoter region of the optineurin gene, associated with a glaucoma
XX XX phenotype), detecting a SNP sequence variation in a sample containing
XX XX DNA, detecting the presence of an optineurin promoter sequence variation
XX XX in a sample containing DNA, determining the presence or increased
XX XX susceptibility to glaucoma or to a progressive ocular hypertensive
XX XX disorder resulting in loss of visual field in a patient (or the severity
XX XX or progression of glaucoma in a patient, comprising providing
XX XX amplification reaction primers that direct amplification of a selected
XX XX nucleic acid region containing the variation within the optineurin
XX XX promoter and amplifying the DNA) and detecting a polymorphism (comprising
XX XX obtaining a sample containing human genomic DNA, providing a nucleic acid
XX XX capable of detecting a SNP located within an optineurin promoter, and
XX XX detecting the polymorphism). The invention is used to diagnose and
XX XX prognose glaucoma and also to treat glaucoma related disorders. The
XX XX present sequence is the optineurin gene.
XX XX
XX XX Sequence 46951 BP; 12703 A; 10108 C; 10051 G; 14089 T; 0 U; 0 Other;
SQ
XX XX
XX XX Query Match 38.2%; Score 25.6; DB 10; Length 46951;
XX XX Best Local Similarity 66.1%; Pred. No. 39;
XX XX Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
XX XX
XX XX 9 TCGAATTAAGAAGCTGCCACAGTTGGTAAGTTAGGTTTAAGACAGTACAG 64
XX XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX XX Db 3290 TGGGATTACAGTGTGAGCCACCGTCCCACTATAGGTTTAAACAGTAAAG 3345
XX XX
XX XX RESULT 14
XX XX ADA30717/c
XX XX ID ADA30717 standard; DNA; 1284 BP.
```

```
XX XX
XX XX AC ADA30717;
XX XX DT 20-NOV-2003 (first entry)
XX XX DE DNA encoding Acinetobacter baumannii protein #2004.
XX XX KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX XX vaccine; plant biocontrol agent.
XX XX KW Acinetobacter baumannii.
XX XX OS
XX XX US5652958-B1.
XX XX PN
XX XX 13-MAY-2003.
XX XX PD
XX XX 04-JUN-1999; 99US-00328352.
XX XX PF
XX XX 09-JUN-1998; 98US-0088701P.
XX XX PR
XX XX (GENO-) GENOME THERAPEUTICS CORP.
XX XX PA
XX XX Breton G, Bush D;
XX XX PI
XX XX
XX XX WPI; 2003-576092/54.
XX XX DR P-PSDB; ADA34643.
XX XX
XX XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX XX for diagnosing a bacterial disease, as components of antibacterial
XX XX vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX XX plants.
XX XX Example; SEQ ID NO 2004; 328pp; English.
XX XX
XX XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX XX CC The A. baumannii nucleic acids and polypeptides are useful as reagents
XX XX CC for diagnosing a bacterial disease, as components of antibacterial
XX XX CC vaccines, as targets for antibacterial drugs, to detect the presence of
XX XX CC A. baumannii and other Acinetobacter species, in a sample, in screening
XX XX CC compounds for the ability to interfere with the A. baumannii life cycle
XX XX CC or to inhibit A. baumannii infection, and as biocontrol agents for
XX XX CC plants. The present sequence represents DNA encoding an A. baumannii
XX XX CC protein.
XX XX
XX XX SQ Sequence 1284 BP; 368 A; 240 C; 311 G; 365 T; 0 U; 0 Other;
XX XX
XX XX Query Match 37.9%; Score 25.4; DB 9; Length 1284;
XX XX Best Local Similarity 61.2%; Pred. No. 20;
XX XX Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
XX XX
XX XX 1 CACAGCTTGGAAATTAAGTCTGCCACAGTTGTAAGTTAGGTTTAAGCAGTA 60
XX XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX XX Db 668 CAAACCTTGAACAGCACTTTGGCCTTCAATTGTGATTTTTCAGCAACTT 629
XX XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX XX QY 61 CAAAGATC 67
XX XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX XX Db 628 CACGAC 622
XX XX
XX XX RESULT 15
XX XX ADP49338/c
XX XX ID ADP49338 standard; DNA; 80000 BP.
XX XX AC
XX XX ADP49338;
XX XX
XX XX 26-AUG-2004 (first entry)
XX XX DE Human B-cell chronic lymphocytic leukaemia associated genomic DNA #2.
XX XX KW ds; gene; human; B-cell chronic lymphocytic leukaemia; B-CLL; cytostatic;
XX XX AMB1; cancer; chromosome 12q21-22.
XX XX OS Homo sapiens.
```

```

XX WO2004046376-A2.
PN
XX
PD 03-JUN-2004.
XX
PF 19-NOV-2003; 2003WO-DK000794.
XX
PR 19-NOV-2002; 2002DK-00001792.
XX
PA (HSRI-) HS RIGSHOSPITAL ET.
XX (UYAR-) UNIV ARHUS.
PI
XX Hertz AMB, Leftere H, Kjeme J;
XX WPI, 2004-449754/42.
XX
PT Establishing diagnosis or prognosis of subtype of B-cell chronic
PT lymphocytic leukemia in individual, involves detecting presence or
PT absence of expression product transcribed from human chromosome 12q21-22,
PT in biological sample.
XX
PS Claim 37, Page 73-95; 11pp; English.
XX
CC The present invention relates to a method of establishing diagnosis or
CC prognosis of a subtype of B-cell chronic lymphocytic leukemia (B-CLL) in
CC individual, which involves detecting the presence or absence of an
CC expression product in a biological sample isolated from the individual,
CC where the expression product comprises a nucleotide sequence transcribed
CC from human chromosome 12q21-22. The sequence is designated AMB1. The
CC method is useful for establishing a diagnosis of a subtype of B-CLL in an
CC individual. The sequences can be used for the treatment of cancer, and
CC for the preparation of a medicament for the treatment of cancer. The
CC cancer is B-CLL. The present sequence is a coding sequence from the human
CC chromosome 12q21-22 which is associated with B-CLL.
XX
SQ Sequence 80000 BP; 24867 A; 15522 C; 15290 G; 24321 T; 0 U; 0 Other;
Query Match 37.0%; Score 24.8; DB 12; Length 80000;
Best Local Similarity 72.7%; Pred. No. 90;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 1 CACAACTTGAAATTAAGTCTGCCCAACAGTTGTAATT 44
Db 72091 CAAAATATTTAATAAATCTGACTGAACATCTGTAATT 72048

```

Search completed: February 8, 2005, 13:57:10
 Job time : 94.8783 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55; Search time 18.0955 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67
Sequence: 1 cacaagctcggaataaag.....gttaagacagctacagatc 67

Scoring table: IDENTITY_MNC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26.6	39.7	9262	4 US-09-566-921-74	Sequence 74, Appl
2	25.4	37.9	1284	4 US-09-328-352-2004	Sequence 2004, Ap
3	24.4	36.4	1389	4 US-09-711-164-174	Sequence 174, App
4	24	35.8	2024	4 US-09-643-657-1	Sequence 1, Appli
5	24	35.8	2024	4 US-09-643-657-2	Sequence 2, Appli
6	23.4	34.9	143	4 US-09-513-999C-18762	Sequence 18762, A
7	23.4	34.9	786431	4 US-09-751-389-3	Sequence 3, Appli
8	23.4	34.9	1830121	4 US-09-557-884-1	Sequence 1, Appli
9	23.4	34.9	1830121	4 US-09-643-990A-1	Sequence 1, Appli
10	23.4	34.9	1830121	4 US-10-329-960-1	Sequence 1, Appli
11	23.2	34.6	3414	1 US-08-764-100-17	Sequence 17, Appl
12	23.2	34.6	3414	1 US-08-764-100-21	Sequence 21, Appl
13	23.2	34.6	4970	1 US-08-764-100-14	Sequence 14, Appl
14	23.2	34.6	4970	1 US-08-764-100-20	Sequence 20, Appl
15	23	34.3	268	4 US-09-513-999C-27534	Sequence 27534, A
16	23	34.3	1435	5 PCT-US95-05922A-1	Sequence 1, Appli
17	23	34.3	1601	4 US-09-023-655-97	Sequence 457, App
18	23	34.3	2580	4 US-08-511-485-7	Sequence 7, Appli
19	23	34.3	2580	4 US-09-201-936-7	Sequence 7, Appli
20	23	34.3	2580	4 US-09-011-356-7	Sequence 7, Appli
21	23	34.3	2580	4 US-09-672-717-222	Sequence 222, App
22	23	34.3	2580	4 US-09-201-932-7	Sequence 7, Appli
23	23	34.3	2589	3 US-08-569-749-1	Sequence 1, Appli
24	23	34.3	2589	5 PCT-US96-12860-1	Sequence 1, Appli
25	23	34.3	3171	2 US-08-868-786-5	Sequence 5, Appli
26	23	34.3	3532	2 US-09-205-204-1	Sequence 1, Appli
27	23	34.3	3732	3 US-09-212-971-7	Sequence 7, Appli

ALIGNMENTS

28	23	34.3	3732	3 US-08-800-929A-7	Sequence 7, Appli
29	23	34.3	3732	3 US-09-617-053A-7	Sequence 7, Appli
30	22.8	34.0	356	4 US-09-640-173-35	Sequence 35, Appl
31	22.8	34.0	356	4 US-09-713-550-35	Sequence 35, Appl
32	22.8	34.0	356	4 US-09-825-294-35	Sequence 35, Appl
33	22.8	34.0	356	4 US-09-970-966-35	Sequence 35, Appl
34	22.8	34.0	892	4 US-09-270-767-2623	Sequence 2623, Ap
35	22.8	34.0	892	4 US-09-270-767-17905	Sequence 17905, A
36	22.8	34.0	1055	4 US-09-270-767-9711	Sequence 9711, Ap
37	22.8	34.0	1055	4 US-09-270-767-24993	Sequence 24993, A
38	22.8	34.0	1249	4 US-09-220-132-4	Sequence 4, Appli
39	22.6	33.7	378	4 US-09-288-143-39	Sequence 39, Appl
40	22.4	33.4	202	4 US-09-513-999C-33953	Sequence 33953, A
41	22.4	33.4	843	4 US-09-248-796A-3158	Sequence 3158, Ap
42	22.4	33.4	957	4 US-09-134-000C-195	Sequence 195, App
43	22.4	33.4	1861	4 US-09-129-668-7	Sequence 7, Appli
44	22.4	33.4	2065	4 US-09-129-668-5	Sequence 5, Appli
45	22.4	33.4	2084	4 US-09-222-938A-9	Sequence 9, Appli

```

RESULT 1
US-09-566-921-74/C
; Sequence 74, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 74
; LENGTH: 9262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No. 6682888 246655.92
; LOCATION: 1941-1942, 1952-1953, 1956, 1958, 1963, 6600-6601
; OTHER INFORMATION: a, t, c, g, or other
US-09-566-921-74

Query Match      39.7%; Score 26.6; DB 4; Length 9262;
Best Local Similarity 63.1%; Pred. No. 1.3;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY      3 CAAGCTTGAAATTAAGTGTGCCACCACTGTGAGTTAGGGTTAAGACAGTACA 62
Db      5902 CAAATAAGATTAATTAAGTGTGCCCACTGTGATCTTAAGATTAAGTATCA 5843
QY      63 AGATTC 67
Db      5842 TGATTC 5838

RESULT 2
US-09-328-352-2004/C
; Sequence 2004, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04

```

NUMBER OF SEQ ID NOS: 8252
SEQ ID NO: 2004
LENGTH: 1284
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-2004

Query Match 37.9%; Score 25.4; DB 4; Length 1284;
Best Local Similarity 61.2%; Pred. No. 2.1;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCCAGCAGTTGGTAATTAGGTTTAAAGACAGTA 60
DB 688 CAAAGCTTGAAGCAGCACTTTGCGCTTCAATTGGTAAGTATTTTTCAGCAACTT 629

QY 61 CAAGATC 67
DB 628 CACGACC 622

RESULT 3

US-09-711-164-174/C
Sequence 174, Application US/09711164
Patent No. 6589738
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Allyn
APPLICANT: Ohlsen, Karl
APPLICANT: Zvekind, Judith
TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
FILE REFERENCE: ELITRA.008A
CURRENT APPLICATION NUMBER: US/09/711.164
PRIOR FILING DATE: 2000-11-09
PRIORITY FILING DATE: US 60/164415
NUMBER OF SEQ ID NOS: 469
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 174
LENGTH: 1389
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1389)
US-09-711-164-174

Query Match 36.4%; Score 24.4; DB 4; Length 1389;
Best Local Similarity 60.6%; Pred. No. 5.3;
Matches 40; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCCAGCAGTTGGTAATTAGGTTTAAAGACAGTA 60
DB 1230 CATCAGCCCGAAGCCGAGATGACCCACAGCGGTTGTAATCGTTCGGTTTACAGACCA 1171

QY 61 CAAGAT 66
DB 1170 GATGAT 1165

RESULT 4

US-09-643-657-1
Sequence 1, Application US/09643657
Patent No. 6642024
GENERAL INFORMATION:
APPLICANT: Diane Pennica
TITLE OF INVENTION: GUANYLATE-BINDING PROTEIN
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,657
FILING DATE: 17-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/015,089A
FILING DATE: 29-Jan-1998
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: P1056
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1896
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2024 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-657-1

Query Match 35.8%; Score 24; DB 4; Length 2024;
Best Local Similarity 64.3%; Pred. No. 8.5;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 AAGCTTGGAAATTAAGTCTGCCCCCAGCAGTTGGTAATTAGGTTTAAAGACAGT 59
DB 1937 AAGCTGGATGAATCAAGCTCCACTTGAATGATATGTTAGGCAAGT 1992

RESULT 5

US-09-643-657-2/C
Sequence 2, Application US/09643657
Patent No. 6642024
GENERAL INFORMATION:
APPLICANT: Diane Pennica
TITLE OF INVENTION: GUANYLATE-BINDING PROTEIN
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,657
FILING DATE: 17-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/015,089A
FILING DATE: 29-Jan-1998
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: P1056
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1896
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 2024 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-643-657-2

Query Match 35.8%; Score 24; DB 4; Length 2024;
Best Local Similarity 64.3%; Pred. No. 8.5;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 AACCTGGAATAAAGCTGCCCCCAGTGTGTAAGGTTTAAGACAGT 59
DB 88 AAGCTGGTATGATCCAGCTCCACTTATGATATGTTAGCAAGT 33

RESULT 6
US-09-513-999C-18762/C
Sequence 18762, Application US/09513999C
Patent No. 6783961

GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961

FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513.999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 18762

LENGTH: 143
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 116
OTHER INFORMATION: k=g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 117
OTHER INFORMATION: r=a or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: 125
OTHER INFORMATION: m=a or c
US-09-513-999C-18762

Query Match 34.9%; Score 23.4; DB 4; Length 143;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGCTGCCCCCAGTGTGTAAGGTTTAAGACAGT 62
DB 91 CAATCTTCTAATATGAGCAATTCGAAGTGTTAATCACTGATATATACATT 32

QY 63 AGATC 67
DB 31 AATTC 27

RESULT 7
US-09-751-389-3/C
Sequence 3, Application US/09751389
Patent No. 6630314
GENERAL INFORMATION:
APPLICANT: GUGELER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL001067
CURRENT APPLICATION NUMBER: US/09/751.389
CURRENT FILING DATE: 2001-01-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 786431

QY 8 TTGGAATAAAGCTGCCCCCAGTGTGTAAGGTTTAAGACAGTACAG 64
DB 62963 TTGGCAATATATAGCTTCCGAAAGGACAGAAATTTCTGTTAAGAGCAATAG 62907

Query Match 34.9%; Score 23.4; DB 4; Length 786431;
Best Local Similarity 63.2%; Pred. No. 80;
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

RESULT 8
US-09-557-884-1/C
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 34.9%; Score 23.4; DB 4; Length 1830121;
Best Local Similarity 60.0%; Pred. No. 93;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAGCTTGGAAATAAAGTGTGCCCCCAGTTGGTAACCTTAGGGTTAAGACAGTA 60
Db 1518906 CATACGTTTACCAATTAATTTGCCCTACACTTGGTTTGTGTGACACACCGTG 1518847

QY 61 CAAGA 65
Db 1518846 TAATA 1518842

RESULT 9

US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 34.9%; Score 23.4; DB 4; Length 1830121;
Best Local Similarity 60.0%; Pred. No. 93;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAGCTTGGAAATAAAGTGTGCCCCCAGTTGGTAACCTTAGGGTTAAGACAGTA 60

Db 1518906 CATACGTTTACCAATTAATTTGCCCTACACTTGGTTTGTGTGACACACCGTG 1518847

QY 61 CAAGA 65

Db 1518846 TAATA 1518842

RESULT 10

US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Patent No. 6742927

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra

Patent No. 6742927
; Title of Invention: Thereof, and Uses Thereof

FILE REFERENCE: PB186P1

CURRENT APPLICATION NUMBER: US/10/329,960

CURRENT FILING DATE: 2003-01-02

PRIOR APPLICATION NUMBER: US 09/643,990

PRIOR FILING DATE: 2000-08-23

PRIOR APPLICATION NUMBER: US 08/487,429

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: US 08/426,787

PRIOR FILING DATE: 1995-04-21

NUMBER OF SEQ ID NOS: 1

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 1830121

TYPE: DNA

ORGANISM: Haemophilus influenzae

FEATURE:

NAME/KEY: misc feature

LOCATION: (4747)..(4747)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (9921)..(9921)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (10150)..(10150)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (29298)..(29298)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (36543)..(36543)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (36551)..(36551)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (40808)..(40810)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (44416)..(44416)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (44975)..(44975)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

NAME/KEY: misc feature

LOCATION: (45553)..(45593)

OTHER INFORMATION: n equals a, t, g or c

FEATURE:

```
/ NAME/KEY: misc_feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51334)..(51334)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51602)..(51602)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (100091)..(100091)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (11924)..(11924)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
```

```
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
```

```
Query Match 34.9%; Score 23.4; DB 4; Length 1830121;
Best Local Similarity 60.0%; Pred. No. 93;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 CACAGCTTGAATTAAGCTGCCCCACCACTTGTACTTACCTTACGCTTTAAGACAGTA 60
DB 1518906 CATACGTTAGCCAAATTAATTGCCCTAACCTTGTTGTGTGTCACACCGTG 1518847
QY 61 CAAGA 65
DB 1518846 TAAATA 1518842
```

```
RESULT 11
US-08-764-100-17/c
; Sequence 17, Application US/08764100
```

Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 3414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-17

Query Match 34.6%; Score 23.2; DB 1; Length 3414;
Best Local Similarity 70.5%; Pred. No. 20;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATTAAGTCTGCCCCCAGCTTGTAAGGTTAG 54
DB 304 GCAATTAATCTGACACATCAGTTGGAGCTTTGGACTATG 261

RESULT 12
US-08-764-100-21
Sequence 21, Application US/08764100
Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-21

Query Match 34.6%; Score 23.2; DB 1; Length 3414;
Best Local Similarity 70.5%; Pred. No. 20;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 GAATTAAGTCTGCCCCCAGCTTGTAAGGTTAG 54
DB 311 GCAATTAATCTGACACATCAGTTGGAGCTTTGGACTATG 3154

RESULT 13
US-08-764-100-14/C
Sequence 14, Application US/08764100
Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800

```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: US/08/214,064
3 FILING DATE:
4 APPLICATION NUMBER: US 08/032,235
5 FILING DATE: 17-MAR-1993
6 APPLICATION NUMBER: GB 9206016.9
7 FILING DATE: 19-MAR-1992
8 ATTORNEY/AGENT INFORMATION:
9 NAME: No. 577370019, Allen E.
10 REGISTRATION NUMBER: 34,490
11 REFERENCE/DOCKET NUMBER: 137-1061
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (415) 354-3592
14 TELEFAX: (415) 857-1125
15 INFORMATION FOR SEQ ID NO: 14:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 4970 base pairs
18 TYPE: nucleic acid
19 STRANDEDNESS: single
20 TOPOLOGY: linear
21 US-08-764-100-14
22
23 Query Match 34.6%; Score 23.2; DB 1; Length 4970;
24 Best Local Similarity 70.5%; Pred. No. 23;
25 Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
26
27 Oy 11 GAATATAAGTCTGCCACCACCGTGTGTAACCTTAGGCGTTAAG 54
28 1774 GCAATATAATTCTGCACATCATCGTTTGAGCGTTTGGAGCATATG 1731
29
30 RESULT 14
31 US-08-764-100-20
32 Sequence 20, Application US/08764100
33 Patent No. 5773700
34 GENERAL INFORMATION:
35 APPLICANT: van Grinsven J., Martinus Q.
36 APPLICANT: De Haan, Petrus T.
37 APPLICANT: Gielen L., Johannes J.
38 APPLICANT: Peters, Dirk
39 APPLICANT: Goldbach, Robert W.
40 TITLE OF INVENTION: Improvements in or Relating to Organic
41 TITLE OF INVENTION: Compounds
42 NUMBER OF SEQUENCES: 27
43 CORRESPONDENCE ADDRESS:
44 ADDRESSEE: Sandoz Agro, Inc
45 STREET: 975 California Avenue
46 CITY: Palo Alto
47 STATE: CA
48 COUNTRY: USA
49 ZIP: 94304
50 COMPUTER READABLE FORM:
51 MEDIUM TYPE: Floppy disk
52 COMPUTER: IBM PC Compatible
53 OPERATING SYSTEM: PC-DOS/MS-DOS
54 SOFTWARE: Patent Release #1.0, Version #1.25
55 CURRENT APPLICATION DATA:
56 APPLICATION NUMBER: US/08/764,100
57 FILING DATE: 06-DEC-1996
58 CLASSIFICATION: 800
59 PRIOR APPLICATION DATA:
60 APPLICATION NUMBER: US/08/214,064
61 FILING DATE:
62 APPLICATION NUMBER: US 08/032,235
63 FILING DATE: 17-MAR-1993
64 APPLICATION NUMBER: GB 9206016.9
65 FILING DATE: 19-MAR-1992
66 ATTORNEY/AGENT INFORMATION:
67 NAME: No. 577370019, Allen E.
68 REGISTRATION NUMBER: 34,490
69 REFERENCE/DOCKET NUMBER: 137-1061
70 TELECOMMUNICATION INFORMATION:
71 TELEPHONE: (415) 354-3592
72 TELEFAX: (415) 857-1125
73 INFORMATION FOR SEQ ID NO: 14:
74 SEQUENCE CHARACTERISTICS:
75 LENGTH: 4970 base pairs
76 TYPE: nucleic acid
77 STRANDEDNESS: single
78 TOPOLOGY: linear
79 US-08-764-100-14

```

```

: TELEFAX: (415) 857-1125
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 4970 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: US-08-764-100-20

Query Match      34.6%; Score 23.2; DB 1; Length 4970;
Best Local Similarity 70.5%; Pred. No. 23;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

DQ      11  GAAATAAAGCTGCCCCCAGCAGTTGGTAACCTTTAGGGTTAG 54
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3197  GCAATATATATTCTGACACATCATGATTGGAGCTTTGGGACTATG 3240

RESULT 15
US-09-513-999C-27534/c
: Sequence 27534, Application US/09513999C
: Patent No. 6783961
: GENERAL INFORMATION:
:   APPLICANT: Dumas Milne Edwards, J.B.
:   APPLICANT: Duclert, A.
:   APPLICANT: Giordano, J.Y.
:   TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
:   Patent No. 6783961
:   FILE REFERENCE: 59. US2.REG
:   CURRENT APPLICATION NUMBER: US/09/513,999C
:   CURRENT FILING DATE: 2000-02-24
:   PRIOR APPLICATION NUMBER: US 60/122,487
:   PRIOR FILING DATE: 1999-02-26
:   NUMBER OF SEQ ID NOS: 36681
:   SOFTWARE: Patent.pm
:   SEQ ID NO 27534
:   LENGTH: 268
:   TYPE: DNA
:   ORGANISM: Homo sapiens
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: 250
:     OTHER INFORMATION: n=a, g, c or t
:   FEATURE:
:     NAME/KEY: misc_feature
:     LOCATION: 251
:     OTHER INFORMATION: d=a or g or t
: US-09-513-999C-27534

Query Match      34.3%; Score 23; DB 4; Length 268;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 35; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

DQ      11  GAAATAAAGCTGCCCCCAGCAGTTGGTAACCTTTAGGGTTAGACAGTACAGA 65
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      113  GAAATAAAGTATAGAGCCCATATATTGGGGAATTTGAGTTTAAACAATTCAATA 59

Search completed: February 8, 2005, 22:26:12
Job time : 26.2066 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 90.2823 Seconds
(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67
Sequence: 1 cacaagcttggaataaag.....gttaagacagctacagatc 67

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/prodata/2/pubpna/US09C_NEW_PUB.seq.*
12: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
18: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
19: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35	52.2	600	US-10-027-632-287050	Sequence 287050,
C 2	35	52.2	600	US-10-027-632-287050	Sequence 287050,
C 3	33.8	50.4	600	US-10-027-632-287049	Sequence 287049,
C 4	33.8	50.4	600	US-10-027-632-287049	Sequence 287049,
C 5	33.8	50.4	259202	US-10-723-860-1311	Sequence 1311, Ap
C 6	26.6	39.7	216	US-10-066-543-2458	Sequence 2458, Ap
C 7	26.6	39.7	559	US-10-060-036-3636	Sequence 3636, Ap
C 8	26.6	39.7	614	US-09-969-034-4304	Sequence 4304, Ap
C 9	26.6	39.7	2516	US-10-723-860-5715	Sequence 5715, Ap
C 10	26.6	39.7	18	US-10-723-860-5715	Sequence 5715, Ap
C 11	26.6	39.7	9277	US-10-084-817-95	Sequence 7293, Ap
C 12	26.4	39.4	617	US-10-027-632-282653	Sequence 282653,

13	26.4	39.4	617	13	US-10-027-632-282654	Sequence 282654,
14	26.4	39.4	617	15	US-10-027-632-282653	Sequence 282653,
15	26.4	39.4	617	15	US-10-027-632-282654	Sequence 282654,
16	26.4	39.4	89856	17	US-10-322-281-79	Sequence 79, Appl
17	26.2	39.1	47804	15	US-10-017-161-717	Sequence 717, Appl
18	26.2	39.1	47804	15	US-10-292-796-629	Sequence 629, Appl
19	26	38.8	1467	16	US-10-425-114-10667	Sequence 10667, A
C 20	26	38.8	3170	16	US-10-424-599-105269	Sequence 105269,
C 21	26	38.8	392112	18	US-10-812-232-3	Sequence 3, Appl1
C 22	25.6	38.2	455	10	US-09-918-995-295	Sequence 295, Appl
C 23	25.6	38.2	2100	15	US-10-094-749-824	Sequence 824, Appl
C 24	25.6	38.2	5054	15	US-10-091-281-1	Sequence 1, Appl1
C 25	25.6	38.2	44728	17	US-10-367-094-23	Sequence 23, Appl
C 26	25.6	38.2	46951	15	US-10-091-281-2	Sequence 2, Appl1
C 27	25.4	37.9	30752	13	US-10-087-192-745	Sequence 745, Appl
C 28	25	37.3	413	16	US-10-424-599-76447	Sequence 76447, A
C 29	24.8	37.0	2207	17	US-10-437-963-10374	Sequence 10374, A
C 30	24.6	36.7	460	16	US-10-424-599-83824	Sequence 83824, A
C 31	24.6	36.7	470	9	US-09-864-761-1997	Sequence 1997, Ap
C 32	24.6	36.7	515	15	US-10-029-366-11997	Sequence 11997, A
C 33	24.6	36.7	879	16	US-10-296-115-595	Sequence 595, Appl
C 34	24.6	36.7	1147	13	US-10-027-632-118119	Sequence 118119,
C 35	24.6	36.7	1147	15	US-10-027-632-118119	Sequence 118119,
C 36	24.6	36.7	49031	17	US-10-322-281-523	Sequence 523, Appl
C 37	24.6	36.7	65359	9	US-09-804-472-3	Sequence 3, Appl1
C 38	24.4	36.4	487	9	US-09-292-758-84	Sequence 84, Appl
C 39	24.4	36.4	1389	15	US-10-287-274-114	Sequence 174, Appl
C 40	24.4	36.4	1389	15	US-10-369-493-47263	Sequence 47263, A
C 41	24.4	36.4	1389	16	US-10-282-122A-6469	Sequence 6469, Ap
C 42	24.4	36.4	27893	15	US-10-017-161-757	Sequence 757, Appl
C 43	24.2	36.1	376	16	US-10-424-599-72155	Sequence 72155, A
C 44	24.2	36.1	383	10	US-09-918-995-18275	Sequence 18275, A
C 45	24.2	36.1	1242	18	US-10-723-860-6348	Sequence 6348, Ap

ALIGNMENTS

RESULT 1
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050
Query Match 52.2%; Score 35; DB 13; Length 600;
Best Local Similarity 79.1%; Pred. No. 0.0044;

Matches 53; Conservative 1; Mismatches 11; Indels 2; Gaps 1;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCACCAAGTTGGTAACCTTTAGGGTTTAAGACAGTA 60

Db 163 CACAAGCTTGGACATTAATAATCTGCYC--ATAGTTGGTGAATTAAGGTTTAAACAGTA 106

QY 61 CAAGATC 67

Db 105 TGAGATC 99

RESULT 2

US-10-027-632-287050/c

Sequence 287050, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287050

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287050

Query Match 52.2%; Score 35; DB 15; Length 600;

Best Local Similarity 79.1%; Pred. No. 0.0044;

Matches 53; Conservative 1; Mismatches 11; Indels 2; Gaps 1;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCACCAAGTTGGTAACCTTTAGGGTTTAAGACAGTA 60

Db 163 CACAAGCTTGGACATTAATAATCTGCYC--ATAGTTGGTGAATTAAGGTTTAAACAGTA 106

QY 61 CAAGATC 67

Db 105 TGAGATC 99

RESULT 3

US-10-027-632-287049/c

Sequence 287049, Application US/10027632

Publication No. US20020198371A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287049

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287049

Query Match 50.4%; Score 33.8; DB 13; Length 600;

Best Local Similarity 79.1%; Pred. No. 0.013; Mismatches 12; Indels 2; Gaps 1;

Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCACCAAGTTGGTAACCTTTAGGGTTTAAGACAGTA 60

Db 163 CACAAGCTTGGACATTAATAATCTGCYC--ATAGTTGGTGAATTAAGGTTTAAACAGTA 106

QY 61 CAAGATC 67

Db 105 TGAGATC 99

RESULT 4

US-10-027-632-287049/c

Sequence 287049, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287049

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287049

Query Match 50.4%; Score 33.8; DB 15; Length 600;

Best Local Similarity 79.1%; Pred. No. 0.013; Mismatches 12; Indels 2; Gaps 1;

Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 1 CACAAGCTTGGAAATTAAGTCTGCCCCACCAAGTTGGTAACCTTTAGGGTTTAAGACAGTA 60

Db 163 CACAAGCTTGGACATTAATAATCTGCYC--ATAGTTGGTGAATTAAGGTTTAAACAGTA 106

QY 61 CAAGATC 67

Db 105 TGAGATC 99

Db 105 TGAGATC 99

RESULT 5
US-10-723-860-1311; Sequence 1311, Application US/10723860
; Publication No. US20040253606A1

; GENERAL INFORMATION:

; APPLICANT: Aziz, Natsaba

; APPLICANT: Ginsburg, Wendy M.

; APPLICANT: Zlotnick, Albert

; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators

; FILE REFERENCE: 05882.0193.NPUS01

; CURRENT APPLICATION NUMBER: US/10/723,860

; CURRENT FILING DATE: 2003-11-26

; PRIOR APPLICATION NUMBER: 60/429,739

; PRIOR FILING DATE: 2002-11-26

; NUMBER OF SEQ ID NOS: 8393

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1311

; LENGTH: 259202

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-723-860-1311

Query Match 50.4%; Score 33.8; DB 18; Length 259202;
Best Local Similarity 79.1%; Pred. No. 0.089;
Matches 53; Conservative 0; Mismatches 12; Indels 2; Gaps 1;

QY 1 CACAGCTTGGAATAAAGCTGCCCCACAGTTGGTAACCTTAGGGTTTAAGACAGTA 60

Db 65966 CACAACTTGACATTAATTAATCTGCTC-ATAGTGTGTAATTAAGGTTTAACAGTA 70023

QY 61 CAAGATC 67

Db 70024 TGAGATC 70030

RESULT 6

US-10-066-543-2458/c

; Sequence 2458, Application US/10066543

; Publication No. US20030087818A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu

; APPLICANT: Pyle, Ruth A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Indrtae, Carol Joseph

; APPLICANT: Lodes, Michael J.

; APPLICANT: Secrist, Heather

; APPLICANT: Carter, Derrick

; APPLICANT: Fanger, Gary R.

; APPLICANT: Smith, Carole L.

; APPLICANT: Durham, Margerita

; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543

; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2458

; LENGTH: 216

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 150_202

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2458
Query Match 39.7%; Score 26.6; DB 14; Length 216;Best Local Similarity 63.1%; Pred. No. 5.3;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGGAATAAAGCTGCCCCACAGTTGGTAACCTTAGGGTTTAAGACAGTACA 62

Db 81 CAAATATGATTAATTAATTAATGTTGGCCCCCAGTTGTTATCTCTAAGATTAAGACTAATCAA 22

QY 63 AGATC 67

Db 21 TGATC 17

RESULT 7

US-10-060-036-3636/c

; Sequence 3636, Application US/10060036

; Publication No. US20030073144A1

; GENERAL INFORMATION:

; APPLICANT: Benson, Darin R.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Persing, David H.

; APPLICANT: Hepler, William T.

; APPLICANT: Jiang, Yugu

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER

; FILE REFERENCE: 210121.566

; CURRENT APPLICATION NUMBER: US/10/060,036

; CURRENT FILING DATE: 2002-01-30

; NUMBER OF SEQ ID NOS: 4560

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3636

; LENGTH: 559

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-060-036-3636

Query Match 39.7%; Score 26.6; DB 14; Length 559;
Best Local Similarity 63.1%; Pred. No. 7.2;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CACAGCTTGGAATAAAGCTGCCCCACAGTTGGTAACCTTAGGGTTTAAGACAGTACA 62

Db 293 CAAATATGATTAATTAATTAATGTTGGCCCCCAGTTGTTATCTCTAAGATTAAGACTAATCAA 234

QY 63 AGATC 67

Db 233 TGATC 229

RESULT 8

US-09-969-034-4304

; Sequence 4304, Application US/09969034

; Publication No. US20040110668A1

; GENERAL INFORMATION:

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Ascle, Jon H.

; APPLICANT: Carroll, Eddie III

; APPLICANT: Catino, Theodore J.

; APPLICANT: Dwivedi, Pooranma

; APPLICANT: Molino, Gary A.

; APPLICANT: Thiagalingam, Arunthathi

; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; TITLE OF INVENTION: Expressed in Cancer Tissue

; FILE REFERENCE: 1657/1032

; CURRENT APPLICATION NUMBER: US/09/969,034

; CURRENT FILING DATE: 2001-10-02

; PRIOR APPLICATION NUMBER: 60/237,271

; PRIOR FILING DATE: 2000-02-10

; NUMBER OF SEQ ID NOS: 4494

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4304

; LENGTH: 614

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 225, 481, 514, 524, 599
OTHER INFORMATION: n = A, T, C or G
US-09-969-034-4304

Query Match 39.7%; Score 26.6; DB 11; Length 614;
Best Local Similarity 63.1%; Pred. No. 7.4;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62
DB 261 CAATAATGATTAATAAATGTTGTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 320

QY 63 AGATC 67
DB 321 TGATC 325

RESULT 9

US-10-723-860-5715/c
Sequence 5715, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05862, 0193, NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
PRIOR FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5715
LENGTH: 2516
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (475)..(489)
OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5715

Query Match 39.7%; Score 26.6; DB 18; Length 2516;
Best Local Similarity 63.1%; Pred. No. 12;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62
DB 1665 CAATAATGATTAATAAATGTTGTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 1606

QY 63 AGATC 67
DB 1605 TGATC 1601

RESULT 10

US-10-723-860-7293/c
Sequence 7293, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882, 0193, NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860

CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 7293
LENGTH: 2516
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (475)..(489)
OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7293

Query Match 39.7%; Score 26.6; DB 18; Length 2516;
Best Local Similarity 63.1%; Pred. No. 12;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62
DB 1665 CAATAATGATTAATAAATGTTGTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 1606

QY 63 AGATC 67
DB 1605 TGATC 1601

RESULT 11

US-10-084-817-95/c
Sequence 95, Application US/10084817
Publication No. US20030119009A1
GENERAL INFORMATION:
APPLICANT: Susan Stuart
APPLICANT: Jed G. Nuchtern
APPLICANT: Sharon E. Plon
APPLICANT: Jason M. Shohet
TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
FILE REFERENCE: PA-0046 US
CURRENT APPLICATION NUMBER: US/10/084,817
PRIOR FILING DATE: 2002-02-25
PRIOR APPLICATION NUMBER: 60/270,784
PRIOR FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 365
SOFTWARE: PERL Program
SEQ ID NO 95
LENGTH: 9277
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030119009A1 1384695.102
NAME/KEY: unsure
LOCATION: 1941-1942, 1952-1953, 1956, 1958, 1963, 6600-6601, 9271
OTHER INFORMATION: a, t, c, g, or other
US-10-084-817-95

Query Match 39.7%; Score 26.6; DB 15; Length 9277;
Best Local Similarity 63.1%; Pred. No. 18;
Matches 41; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGTCTGCCCCCAGCTTGTAACCTTTAGGTTTAAGACAGTACA 62
DB 5902 CAATAATGATTAATAAATGTTGTGCCCCCAGTTGTTATCTTAAGATTAAGATTAATCAA 5843

QY 63 AGATC 67
DB 5842 TGATC 5838

RESULT 12

US-10-027-632-282653
Sequence 282653, Application US/10027632

```
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 282653
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-282653
```

Query Match 39.4%; Score 26.4; DB 13; Length 617;

Best Local Similarity 65.0%; Pred. No. 8.9; Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
QY 8 TTGGAATTAAGTCTGCCCCCAGTGTGTAAGGTTTAAAGACAGTCAAGATC 67
Db 184 TTGGAGTAAACAGTCTACCCCAAACTTAGTCTCATGTATTAACATTCAAAAAAC 243
```

```
RESULT 13
US-10-027-632-282654
; Sequence 282654, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 282654
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-282654
```

Query Match 39.4%; Score 26.4; DB 13; Length 617;

Best Local Similarity 65.0%; Pred. No. 8.9; Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
QY 8 TTGGAATTAAGTCTGCCCCCAGTGTGTAAGGTTTAAAGACAGTCAAGATC 67
Db 184 TTGGAGTAAACAGTCTACCCCAAACTTAGTCTCATGTATTAACATTCAAAAAAC 243
```

```
RESULT 14
US-10-027-632-282653
; Sequence 282653, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 282653
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-282653
```

Query Match 39.4%; Score 26.4; DB 15; Length 617; Best Local Similarity 65.0%; Pred. No. 8.9; Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
QY 8 TTGGAATTAAGTCTGCCCCCAGTGTGTAAGGTTTAAAGACAGTCAAGATC 67
Db 184 TTGGAGTAAACAGTCTACCCCAAACTTAGTCTCATGTATTAACATTCAAAAAAC 243
```

```
RESULT 15
US-10-027-632-282654
; Sequence 282654, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
```

; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 282654
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-282654

Query Match 39.4%; Score 26.4; DB 15; Length 617;
Best Local Similarity 65.0%; Pred. No. 8.9;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 8 TTGGAAATTAAGTCTGCCCCAGTTGTTAACTTTAGGGTTTAAAGACAGTACAAGATC 67
Db 184 TTGGAGTAAAGTCTGCCCCAGTTGTTAACTTTAGGGTTTAAAGACAGTACAAGATC 243

Search completed: February 9, 2005, 06:51:04
Job time : 93.2823 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 798.033 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-8

Perfect score: 67
Sequence: 1 cacaagcttgcgaataaag.....gtttaagacagcacaagatc 67

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	31	46.3	513	8	AZ725508 RPCI-24-1
C 2	31	46.3	544	8	BH118642 RPCI-24-3
C 3	30.2	45.1	569	2	BF054120 EST439350
C 4	30.2	45.1	703	5	BO509607 EST617022
C 5	29.2	43.6	802	9	CL315981 ZMMBB0050
C 6	28.8	43.0	670	9	CE826965 tigr-gsa-
C 7	28.8	42.7	584	2	BE463065 EST354360
C 8	28.2	42.1	630	2	BB604453 BB604453
C 9	28.2	42.1	708	8	AZ263719 RPCI-23-1
C 10	28.2	42.1	729	9	AG505646 Mus muscu
C 11	27.8	41.5	232	2	BB214296 BB214296
C 12	27.8	41.5	469	8	AZ617508 1M0448N17
C 13	27.6	41.2	421	2	BB786603 BB786603
C 14	27.4	40.9	574	7	CN667917 A0860C04-
C 15	27.4	40.9	587	8	AQ410748 HS_5067_A
C 16	27.4	40.9	698	9	AG297816 Mus muscu
C 17	27.4	40.9	726	9	AG295587 Mus muscu
C 18	27.2	40.6	598	7	CO379630 FRA2964 S
C 19	27.2	40.6	1002	1	AL576285 AT576285
C 20	27	40.3	408	8	AO635311 RPCI-11-4
C 21	27	40.3	772	1	CL375739 ZMMBB0039
C 22	26.8	40.0	206	1	AV255261 AV255261
C 23	26.8	40.0	351	1	AJ559892 AJ559892
C 24	26.8	40.0	413	4	BG958108 CMO-CT080

C 25	26.8	40.0	672	9	BX147161 BX147161
C 26	26.8	40.0	701	9	BX185444 BX185444
C 27	26.6	39.7	294	7	T04932 T04932
C 28	26.6	39.7	331	1	AA330839 AA330839
C 29	26.6	39.7	342	1	AA262024 AA262024
C 30	26.6	39.7	443	1	AA442029 AA442029
C 31	26.6	39.7	488	6	CA333757 CA333757
C 32	26.6	39.7	489	8	BZ62614 BZ62614
C 33	26.6	39.7	498	1	AA176275 AA176275
C 34	26.6	39.7	499	1	AA176332 AA176332
C 35	26.6	39.7	540	2	AW956641 AW956641
C 36	26.6	39.7	572	9	CR333642 CR333642
C 37	26.6	39.7	584	5	BU072790 BU072790
C 38	26.6	39.7	635	6	CA948269 CA948269
C 39	26.6	39.7	641	6	CB127044 CB127044
C 40	26.6	39.7	699	4	BG623734 BG623734
C 41	26.6	39.7	713	1	AV717150 AV717150
C 42	26.6	39.7	737	5	BX411376 BX411376
C 43	26.6	39.7	777	5	BM724386 BM724386
C 44	26.6	39.7	797	5	BX473434 BX473434
C 45	26.6	39.7	832	9	CR293936 CR293936

ALIGNMENTS

RESULT 1
AZ725508 513 bp DNA linear GSS 24-JAN-2001
RPCI-24-109J1.TV RPCI-24 Mus musculus genomic clone RPCI-24-109J1,
DEFINITION genomic survey sequence.

ACCESSION AZ725508
VERSION AZ725508.1 GI:12472231
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 513)
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintre,B., Levins,M., Tsagaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregiorgis,E., Russell,D., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.tigr.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Place: 109 row: J column: 1
Seq primer: T7
Class: BAC ends.

FEATURES

source Location/Qualifiers
1..513
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-109J1"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTRABAC1, Site_1: BamHI, Site_2: BamHI, RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTRABAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J

ORIGIN

DNA."

Query Match 46.3%; Score 31; DB 8; Length 513;
 Best Local Similarity 76.1%; Pred. No. 2.3;
 Matches 51; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 2 ACAAGCTTGAATAAAGTCTGCCACACAGTTGTAAGTTT-AGGGTTTAAGACAGTA 60
 DB 463 ATATATCTTGAGACTAGAAATCTCTCCCATGTTGTAATTTTCACGTGTGAACAGATG 404
 QY 61 CAAGATC 67
 DB 403 CAAGATC 397

RESULT 2

BH118642/c 544 bp DNA linear GSS 19-JUN-2001
 LOCUS RPCI-24-359A2.TJ RPCI-24 Mus musculus genomic clone RPCI-24-359A2,
 DEFINITION genomic survey sequence.
 ACCESSION BH118642
 VERSION BH118642.1 GI:14961134
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 544)
 Authors Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akincet, B., Levins, M.,
 Tregeve, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
 Russell, D., de Jong, P. and Fraser, C. M.
 Mouse BAC End Sequences from Library RPCI-24

TITLE Unpublished (1999)
 JOURNAL Other GSSes: RPCI-24-359A2.TV
 COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cno.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 359 row: A column: 2
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..544

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="CS7BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-359A2"
 /sex="Male"
 /cell_type="spleen/brain"
 /clone_id="RPCI-24"
 /note="Vector: pTRABAC1, Site 1: BamHI, Site 2: BamHI,
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTRABAC1 cloning vector at the
 BamHI sites using MboI partially digested male CS7BL/6J
 DNA."

ORIGIN

Query Match 46.3%; Score 31; DB 8; Length 544;
 Best Local Similarity 76.1%; Pred. No. 2.3;
 Matches 51; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 2 ACAAGCTTGAATAAAGTCTGCCACACAGTTGTAAGTTT-AGGGTTTAAGACAGTA 60
 DB 463 ATATATCTTGAGACTAGAAATCTCTCCCATGTTGTAATTTTCACGTGTGAACAGATG 404

DB 484 ATATATCTTGAGACTAGAAATCTCTCCCATGTTGTAATTTTCACGTGTGAACAGATG 425

QY 61 CAAGATC 67
 DB 424 CAAGATC 418

RESULT 3

BF054120/c 569 bp mRNA linear EST 07-MAR-2003
 LOCUS EST338D7 5' sequence, mRNA sequence.
 DEFINITION EST338D7 5' sequence, mRNA sequence.
 ACCESSION BF054120
 VERSION BF054120.1 GI:10808016
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 569)
 Authors van der Hoeven, R.S., Bezzerides, J., Holt, I.E., Liang, F., Cho, J.,
 Uteirback, T., Hansen, C.L., Doan, B., Bougri, O., Buell, C.R.,
 Roming, C.M., Fry, W.E., Tanksley, S.D. and Baker, B.
 Generation of ESTs from potato leaves and petioles

Unpublished (2000)
 TITLE Contact: Robin Buell
 JOURNAL The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT Email: potato-array@tigr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
<http://genome.arizona.edu/orders/>.

FEATURES

source Location/Qualifiers
 1..569
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="EST338D7"
 /tissue_type="leaflets and petioles"
 /dev_stage="8 weeks old plants"
 /lab_host="SOLR"
 /clone_id="potato leaves and petioles"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Tissue was supplied by Dr. Fry (Cornell University).
 Leaflets and petioles were isolated from 8 week old
 greenhouse grown plants. The plants were watered and
 fertilized freely. The tissue was immediately frozen in
 liquid nitrogen."

ORIGIN

Query Match 45.1%; Score 30.2; DB 2; Length 569;
 Best Local Similarity 65.7%; Pred. No. 4.6;
 Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 CACAAGCTTGAATAAAGTCTGCCACACAGTTGTAAGTTT-AGGGTTTAAGACAGTA 60
 DB 544 CACATCTTCATCAATTAATTTCTTAACACACAGTCTTAAGGTAAGGTTTGAATGA 485
 QY 61 CAAGATC 67
 DB 484 CAAGTTC 478

RESULT 4

BF0509607/c 703 bp mRNA linear EST 07-MAR-2003
 LOCUS EST617022 Generation of a set of potato cDNA clones for microarray
 DEFINITION analyses mixed potato tissues Solanum tuberosum cDNA clone STMHDB3
 3' end, mRNA sequence.
 ACCESSION BF0509607
 VERSION BF0509607.2 GI:21925341

KEYWORDS
SOURCE EST.
ORGANISM Solanum tuberosum (potato)

REFERENCE
AUTHORS Solanum tuberosum (potato)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 703)
 Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
 Restrepo, S., Griffiths, H., van der Hoeven, K., Tsai, U. and
 Karymchewa, S.A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 On Jun 10, 2002 this sequence version replaced gi:21368476.
 Other ESTs: EST617020 EST617021

TITLE
JOURNAL The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
 http://genome.arizona.edu/orders/
 Seq primer: T7.

FEATURES
source
 1..703
 Location/Qualifiers
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec or Binjite"
 /db_xref="taxon:4113"
 /clone="STM93"
 /issue_type="mixed tissues"
 /lab_host="SOLR"
 /clone_lib="generation of a set of potato cDNA clones for
 microarray analyses mixed potato tissues"
 /note="Vector: Bluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Combination of untreated and Phytophthora
 infestans-treated libraries of scions, leaves, leaflets,
 axillary buds of stem explants, petioles, germinating
 eyes, tubers, or roots."

ORIGIN
 Query Match 45.1%; Score 30.2; DB 5; Length 703;
 Best Local Similarity 65.7%; Pred. No. 4.8;
 Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY
 1 CACAGCTTGGAATAAAGTCTGCCACCAAGTGTGAATTAGGTTTAAGACAGTA 60
 |||||
 548 CACATGCTTCATCATTAATTTCTTCAACACCAAGTCTTAAGGTTTTCACATGA 489
 |||||

QY
 61 CAGATC 67
 |||||

Db
 488 CAGGTT 482

RESULT 5
CL315981 802 bp DNA linear GSS 03-MAR-2004
LOCUS ZM8BC0502B08 Zea mays genomic clone ZM8BC0502B08 3',
DEFINITION genomic survey sequence.
ACCESSION CL315981
VERSION CL315981.1 GI:44895870
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 802)
 Bharti, A.K., Young, S., Kavchok, S., Keifer, G., Bronzino, A.C.,
 Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J.
 Sequencing of the maize genome at F01R (2003c)
 Unpublished (2003)
 Contact: Bharti, A.K.
 Dr. Joachim Messing's lab

KEYWORDS
SOURCE The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bharti@waksman.rutgers.edu
 Seq primer: SP6
 Class: BAC ends.

REFERENCE
AUTHORS Bharti, A.K., Young, S., Kavchok, S., Keifer, G., Bronzino, A.C.,
 Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J.
 Sequencing of the maize genome at F01R (2003c)
 Unpublished (2003)
 Contact: Bharti, A.K.
 Dr. Joachim Messing's lab

TITLE
JOURNAL The Institute for Genomic Research
 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

FEATURES
source
 1..670
 Location/Qualifiers
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ZM8BC0502B08"
 /lab_host="E. coli DH10B"
 /clone_lib="ZM8BC"
 /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"

ORIGIN
 Query Match 43.6%; Score 29.2; DB 9; Length 802;
 Best Local Similarity 69.0%; Pred. No. 11;
 Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY
 5 AGCTTGGAATAAAGTCTGCCACCAAGTGTGAATTAGGTTTAAGACAGTA 62
 |||||
 530 AGATTGAACCTCAAGTGAACACCCAGGTTAGCTTAAGGTTTAAGTCAACACA 587
 |||||

RESULT 6
CE826965/c 670 bp DNA linear GSS 30-SEP-2003
LOCUS tigr-gss-dog-17000331964321 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE826965
VERSION CE826965.1 GI:37167985
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 1 (bases 1 to 670)
 Kirknes, E.F., Batina, V., Halpern, A.L., Levy, S., Remington, K.,
 Ruesch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and
 Venter, U.C.
 The dog genome: survey sequencing and comparative analysis
 Science 301 (5641), 1898-1903 (2003)
 22875432
 14512627
 Contact: Kirknes EF
 The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

TITLE
JOURNAL The Institute for Genomic Research
 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

DEFINITION The Institute for Genomic Research
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
 Rockville, MD 20850, USA
 Tel: 301-838-0200
 Fax: 301-838-0208
 Email: ekirknes@tigr.org
 Class: shotgun.

ORIGIN
 Query Match 43.0%; Score 28.8; DB 9; Length 670;
 Best Local Similarity 65.6%; Pred. No. 15;
 Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY
 1 CACAGCTTGGAATAAAGTCTGCCACCAAGTGTGAATTAGGTTTAAGACAGTA 60

DB 478 CAAAAGCTGAAATTAATTCATGCTCATCCATGCAAGCTTTATGTTTAAAGATTA 419
 QY 61 CAAG 64
 DB 418 AAAG 415

RESULT 7
 BE463065/c 584 bp mRNA linear EST 18-MAY-2001
 LOCUS EST354360 tomato flower buds 8 mm to pre-anthesis, Cornell
 DEFINITION University Lycopersicon esculentum cDNA clone CTC0C11F24, mRNA
 sequence
 BE463065
 ACCESSION BE463065.1 GI:9508908
 VERSION
 KEYWORDS
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 584)
 van der Hoeven, R.S., Bezzerides, J.L., Matern, A.L., Holt, I.E.,
 Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S.,
 Roming, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and
 Tanksley, S.D.
 Generation of ESTs from tomato flower tissue
 Unpublished (1999)
 CONTACT: CUGI
 CLEMSON UNIVERSITY Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 Location/Qualifiers
 1..584
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cullivar="TA496"
 /db_xref="taxon:4081"
 /clone="CTC0C11F24"
 /tissue_type="flower"
 /dev_stage="buds 8mm-to-preanthesis"
 /clone_lib="tomato flower buds 8 mm to pre-anthesis,
 Cornell University"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; supplier: Tanksley; Flower buds and flowers were
 taken from greenhouse plants (4-8 wks old, TA496). They
 were immediately frozen in liquid nitrogen and then
 size-separated while remaining frozen."

ORIGIN
 Query Match 42.7%; Score 28.6; DB 2; Length 584;
 Best Local Similarity 64.2%; Pred. No. 18;
 Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 QY 1 CACAAAGCTTGAATAAAGCTGCCCCCAAGTGTGTAAGCTTTAGAGACAGTA 60
 DB 543 CACATGCTTCATCATTAATTTCTTCAACACCAAGCTTTAAAGCAGAGTTTGGACATGA 484
 QY 61 CAAGATC 67
 DB 483 CAAGTTC 477

RESULT 8
 BB660453/c 630 bp mRNA linear EST 26-OCT-2001
 LOCUS BB660453 RIKEN full-length enriched, 13 days embryo lung Mus
 DEFINITION musculus cDNA clone D430047118 5', mRNA sequence.
 ACCESSION BB660453
 VERSION BB660453.1 GI:16494274

KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 630)
 REFERENCE
 Arikawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,
 Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,
 Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,
 Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, P.,
 Takeeda, X., Tanaka, T., Toyata, T., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arikawa, T., et al. 2001)
 TITLE
 JOURNAL
 COMMENT
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-reseq@riken.jp, URL: <http://genome.gsc.riken.jp/>
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subcloning of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Ohara, E.,
 Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
 Matsuzawa, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
 and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
 Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Aizawa, K., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I.,
 Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1..630
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="D430047118"
 /tissue_type="lung"
 /dev_stage="13 days embryo"
 /lab_host="MDH10B"
 /clone_lib="RIKEN full-length enriched, 13 days embryo
 lung"
 /note="Site_1: SalI; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGACGGCCGCACTCGAGTGTGTTTGTGTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5',
 GAGAGAGAGATTCGAGTTAATTAATTAATCCGCCGCCGCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified

ORIGIN pbuuescript KS(+) after bulk excision from Lambda FLIC I."

Query Match 42.1%; Score 28.2; DB 2; Length 630;
Best Local Similarity 64.6%; Pred. No. 25;
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATAAAGCTTCCCGACAGTTGTAACCTTAAAGCAGTAC 61
DB 532 AGAACTTTAAATACAACTGCCCCAAAGTAGTACTCTGTTAGTGTCAATTG 473

QY 62 AAGAT 66
DB 472 AAGAT 468

RESULT 9 708 bp DNA linear GSS 26-JUL-2000
A2263719
LOCUS RPCI-23-153M12, TV RPCI-23 Mus musculus genomic clone
DEFINITION RPCI-23-153M12, genomic survey sequence.
ACCESSION A2263719
VERSION A2263719.1 GI:9474235
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Bhatnagar, Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 708)

REFERENCE
AUTHORS Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S.,
Akintet, B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de
Jong, P. and Fraser, C.M. Mammals, 1999

TITLE
JOURNAL Mouse BAC End Sequences from Library RPCI-23
COMMENT Unpublished (1999)
Contact: Shaying Zhao
Department of Bukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buhalo.edu/orderingframe.htm>)
or from Resea ch Genetics (inforesgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ende/mouse/bac_end_intro.html
Plate: 153 row: M column: 12
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers

1..708
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-153M12"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1.
ECORI; Site: 2: EORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EORI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match 42.1%; Score 28.2; DB 8; Length 708;
Best Local Similarity 64.6%; Pred. No. 26;
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATAAAGCTTCCCGACAGTTGTAACCTTAAAGCAGTAC 61
DB 495 AGAACTTTAAATACAACTGCCCCAAAGTAGTACTCTGTTAGTGTCAATTG 436
QY 62 AAGAT 66
DB 435 AAGAT 431

RESULT 10 729 bp DNA linear GSS 04-JUN-2004
AG505646
LOCUS Mus musculus molossinus DNA, clone:MSWg01-407M07.T7, genomic survey
sequence.
ACCESSION AG505646
VERSION AG505646.1 GI:48213059
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus

REFERENCE
AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
TITLE BAC end Sequences of Library MSWg01
JOURNAL Unpublished
2 (bases 1 to 729)
REFERENCE Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
AUTHORS Direct Submission
TITLE Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suenho-chu, Tsukuba, Ibaraki, Japan, Yokohama, Kanagawa 230-0045, Japan
(E-mail: hattori@gsc.riken.jp, URL: <http://hyp.gsc.riken.go.jp/>,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the mouse BAC library MSWg01. For BAC
library availability, please contact Kunya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center.
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp

COMMENT

LIBRARY : T7
Vector : pBAC3.6
R.Site 1 : EORI
R.Site 2 : EORI.
Location/Qualifiers

1..729
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSWg01-407M07.T7"
/sex="male"
/issue_type="mixture of kidney and spleen"
/clone_lib="MSWg01 Mouse Male BAC Library"

ORIGIN

Query Match 42.1%; Score 28.2; DB 9; Length 729;
Best Local Similarity 64.6%; Pred. No. 26;
Matches 42; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 ACAAGCTTGAATAAAGCTTCCCGACAGTTGTAACCTTAAAGCAGTAC 61
DB 592 AGAACTTTAAATACAACTGCCCCAAAGTAGTACTCTGTTAGTGTCAATTG 651

QY 62 AAGAT 66
DB 652 AAGAT 656

RESULT 11

TITLE
Riken Mouse ESRs (Komno, H., et al.)
JOURNAL
Unpublished (2000)
COMMENT
Contact: Yoshihide Hayashizaki

FEATURES	Location/Qualifiers
source	1. .232

```

1. .232
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A530019P19"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10+"
/clone_lib="RIKEN full-length enriched, adult male aorta
and vein"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGAGATCCAGAGGCTCTTTTTTTTTTTTTTNN 3'. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization

```

RESULT	12
AZ617508/c	
LOCUS	AZ617508 469 bp DNA linear GSS 13-DEC-2000
DEFINITION	M00448N1R Mouse 10kb plasmid UGCM library Mus musculus genomic clone UGCM00448N1 R, genomic survey sequence.
ACCESSION	AZ617508
VERSION	AZ617508.1 GI:11739698
KEYWORDS	GSS.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

FEATURES

SOURCE

```

"organism"="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M046N17"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb Plasmid UUCG1 library"
/notes="Vector: PMW429; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-ligated with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMW42 (g14732114.gdb.af129072.1), a copy-number

```

Inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 41.5%; Score 27.8; DB 8; Length 469;
Best Local Similarity 65.1%; Pred. No. 33;
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3 CAAGCTTGAATAAAGTCTGCCCGACGTTGTAACCTTAGGCTTAAGACAGTACA 62

DB 205 CACCTGAGAGATTAAGTCTGCACACCCACACATTAATGCTTGAAGCCAGTAC 146

QY 63 AGA 65

DB 145 AGA 143

RESULT 13

BB786603 421 bp mRNA linear EST 08-JUL-2003
LOCUS BB786603 RIKEN full-length enriched, RCB-0035 WEHI-3 CDNA Mus
DEFINITION musculus cDNA clone G430112H24 3', mRNA sequence.

ACCESSION BB786603
VERSION BB786603.1 GI:16955099

KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 421)

REFERENCE 1 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Konda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku, K., Tanaka, T., Tomaru, A., Toyota, T., Watahiki, A., Yasunishi, A., Yamanaka, M., and Hayashizaki, Y.
Muramatsu, M. and Hayashizaki, Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (AKimura, T., et al. 2001)

JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.jp, URL: http://genome-gsc.riken.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

FEATURES

Source

e mouse tissues.
Location/Qualifiers
1..421
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="G430112H24"
/tissue_type="Blood"
/cell_line="RCB-0035 WEHI-3"
/clone_1ib="RIKEN full-length enriched, RCB-0035 WEHI-3 CDNA"

ORIGIN

Query Match 41.2%; Score 27.6; DB 2; Length 421;
Best Local Similarity 67.2%; Pred. No. 38;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 10 GGAATTAAGTCTGCCCGACGTTGTAACCTTAGGCTTAAGACAGTACAAGTAC 67

DB 317 GGCAAGTAAAGTAAAGTCCCGACAGAGTAACTTAATATGCTAAGAACGTATATTATC 374

RESULT 14

CN667917 574 bp mRNA linear EST 17-MAY-2004
LOCUS CN667917/c A0860C04-5 NIA Mouse E13.5 whole embryo cDNA library (long) Mus
DEFINITION musculus cDNA clone NIA:A0860C04 IMAGE:30762939 5', mRNA sequence.

ACCESSION CN667917
VERSION CN667917.1 GI:47434368

KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 574)

REFERENCE 1 Sharov, A.A., Piao, Y., Maroba, R., Dudekula, D.B., Qian, Y., Vanburen, V., Falco, G., Martin, P.R., Sragg, C.A., Basse, U.C., Wang, Y., Carter, M.G., Hamatani, T., Alpa, K., Akutsu, H., Sharova, L., Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S., Nagaraja, R., Boheler, K.R., Taub, D., Hodges, R.J., Longo, D.L., Schlessinger, D., Keller, J., Klotz, E., Kelsoe, G., Umezawa, A., Vescevi, A.L., Rosant, J., Kunath, T., Hogan, B.L., Curci, A., D'Urso, M., Kelso, J., Hide, W. and Ko, M.S.
Transcriptome analysis of mouse stem cells and early embryos
PLOS Biol. 1 (3), 410-419 (2003)
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: A0860 row: C column: 04
Seq primer: M13 Reverse
High quality sequence stop: 574
POLY(A)-No.

TITLE

JOURNAL

COMMENT

FEATURES

Source

Location/Qualifiers
1..574
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="NIA:A0860C04 IMAGE:30762939"
/tissue_type="whole embryo including extraembryonic tissues at 13.5-days postcoitum"
/dev stage="E13.5"
/lab host="DH10B"
/clone_1ib="NIA Mouse E13.5 whole embryo cDNA library (long)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cdna>).

This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from 1 embryo at 13.5-days postcoitum. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]: 5'-pGACTGTTCTTAAATCGCAGCGCCGCTTTTCTTTTCTTTT-3' from 3ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal1, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal1-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with Sal1 and NotI enzymes and cloned into Sal1/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0kb. The library was constructed by Yulan Piao."

ORIGIN

Query Match

40.9%; Score 27.4; DB 7; Length 574;

Best Local Similarity 65.6%; Pred. No. 48;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 5 AGCTTGGAATTAATAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATCAAG 64

DB 303 AGTTTAAATAAAAGATGATACGAGACTGAGGAAACAGTGTGTTAGGAAAGTACAG 244

QY 65 A 65

DB 243 A 243

RESULT 15

LOCUS

AO410748

DEFINITION

587 bp DNA linear GSS 17-MAR-1999
HS_3067_A1.B11.T7A.RPCL-11 Human Male BAC Library Homo sapiens
genomic clone Plate=643 Col=21 Row=1, genomic survey sequence.

ACCESSION

AO410748

VERSION

AO410748.1 GI:4427941

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 587)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

10449764

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCL-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Research Genetics (info@resgen.com). BAC end Web Server:

http://www.brc.washington.edu

Plate: 643 Row: I Column: 21

Seq primer: TV

Class: BAC ends

High quality sequence atop: 587.

Location/Qualifiers

source

1..587

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="Plate=643 Col=21 Row=1"

/sex="male"

/clone_lib="RPCL-11 Human Male BAC Library"

/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI methylase. Size selected DNA was cloned into the

pBAC3.6 vector at EcoRI sites"

ORIGIN

Query Match

40.9%; Score 27.4; DB 8; Length 587;

Best Local Similarity 69.8%; Pred. No. 49;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 11 GAAATTAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATCAAG 63

DB 522 GAAATTAAGTCTGCCACAGTGTGTAACCTTTAGGGTTTAAGACATCAAG 574

Search completed: February 8, 2005, 22:18:16
Job time : 806.033 secs

FEATURES

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 : Search time 797.321 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143
Sequence: 1 gattctgttcgtcgtcctcc.....gtttaagacagctacaagatc 143

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.4	59.7	5562	9 AB091787	AB091787 Ocolemur
2	85.4	59.7	6451	9 AB091789	AB091789 Bos tauru
3	83.8	58.6	5684	9 AB091785	AB091785 Lemur cat
4	83.4	58.3	5712	4 AB091793	AB091793 Equus cab
5	76.2	53.3	6465	9 AB091781	AB091781 Pan trogl
6	74.6	52.2	8810	9 AY040206	AY040206 Homo sapi
7	74.6	52.2	158142	2 AL357130	AL357130 Homo sapi
8	74.6	52.2	259202	9 AC002366	AC002366 Human XP2
9	74.4	52.0	685	11 BV089285	BV089285 RPAMMSBO
10	74.4	52.0	695	11 BV097603	BV097603 RPAMMSBO
11	72.8	50.9	6442	9 AB091783	AB091783 Salimiri s
12	72.2	50.5	463	10 D8306382	D83064 Mus musculu
13	72.2	50.5	9384	10 AF294397	AF294397 Mus musculu
14	72.2	50.5	95826	10 AL805974	AL805974 Mouse dnA
15	71	49.7	7425	4 AB091791	AB091791 Sus scrofa
16	68.8	48.1	177654	9 AP000918	AP000918 Homo sapi
17	68.8	48.1	200214	9 AC013412	AC013412 Homo sapi
18	68.4	47.8	212886	2 AC093946	AC093946 Rattus no
19	68.4	47.8	234471	2 AC121424	AC121424 Rattus no

C 20	68.2	47.7	363	11 BV089294	BV089294 RPAMMSBO
C 21	68.2	47.7	363	11 BV097602	BV097602 RPAMMSBO
C 22	67.2	47.0	7163	9 AB091782	AB091782 Pan trogl
C 23	67.2	47.0	38765	9 BS000568	BS000568 Pan trogl
C 24	67.2	47.0	177726	9 AC145770	AC145770 Pan trogl
C 25	67.2	47.0	190089	9 BS000576	BS000576 Pan trogl
C 26	66.4	46.4	6931	4 AB091792	AB091792 Sus scrofa
C 27	64.2	44.9	5591	4 AB091794	AB091794 Equus cab
C 28	61	42.7	8004	9 AB091786	AB091786 Lemur cat
C 29	56.8	39.7	7454	9 AB091784	AB091784 Salimiri s
C 30	47.2	33.0	6264	4 AB091790	AB091790 Bos tauru
C 31	42.6	29.8	1935	9 D83730	D83730 Homo sapien
C 32	42.4	29.7	5151	9 AB091788	AB091788 Ocolemur
C 33	41.8	29.2	726	10 S74899	S74899 ameleogenin
C 34	41.8	29.2	727	10 MUSNAMEB	D31769 Mus musculu
C 35	41.8	29.2	765	4 AB032194	AB032194 Equus cab
C 36	41.8	29.2	789	4 AB032193	AB032193 Equus cab
C 37	41.8	29.2	798	10 BC059090	BC059090 Mus muscu
C 38	41.8	29.2	799	10 MUSNAMEA	D31768 Mus musculu
C 39	40.2	28.1	65	6 CQ531502	CQ531502 Sequence
C 40	40.2	28.1	359	10 RRU07054	UR07054 Rattus norv
C 41	40.2	28.1	457	10 RRU060564	U60564 Rattus norv
C 42	40.2	28.1	476	6 AR452534	AR452534 Sequence
C 43	40.2	28.1	753	10 RNU60562	U60562 Rattus norv
C 44	40.2	28.1	780	10 U01245	U01245 Rattus norv
C 45	40.2	28.1	812	10 RNU67130	U67130 Rattus norv

ALIGNMENTS

RESULT 1
AB091787
LOCUS . Ocolemur garnettii AMELX gene for amelogenin, partial cds.
DEFINITION AB091787
ACCESSION AB091787
VERSION AB091787.1 GI:29126027
KEYWORDS
SOURCE
ORGANISM
Ocolemur garnettii (small-eared galago)
Ocolemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Ocolemur.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

REFERENCE
AUTHORS
TITLE
JOURNAL
Iwase, M., Saita, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Soken), Department of Biosystems Science; Shonan Kokuuamura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@mekoryu.wol.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)
Location/Qualifiers

FEATURES

source
1..5562
/organism="Ocolemur garnettii"
/mol_type="genomic DNA"
/db_xref="taxon:30611"
/sex="male"
267..5562
/gene="AMELX"
join(267..321,1401..1412)
/gene="AMELX"
join(1413..1466,3385..3432,4699..4740,4832..4876,5125..5562)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66107.1"

```

/crambl:ac:lon="MCGWILFACILGAAAFEMPLPRPHGHGHGYNFSEVNSYFPGISID
ETFLVULTRPKWYSMTIRHRYPSYSGYEMGMSLHQIILPVTSQOTPOUNHAIOPHNHILM
VAPQOAVPVFQOPMMVPGVGSMSYPTOHQONLRLYPAQOPRQOSIQOPRQOPRQOPRQ
LOPMOPRQOPRQOPRQOPRQVNDIOPLRPLQOPRPLRPIFFMQOPRPLRPLMLDIBLEWADAT
"

```

Query Match	58.6%	Score 93.8;	DB 9;	Length 5684;
Best Local Similarity	77.9%;	Pred. No. 1.4e-15;		
Matches 116; Conservative	0;	Mismatches 27;	Indels 6;	Gaps 1;

Dd		1585
Oy	61 -----GCATTAATATTCAATTTCACAAAGCTTGGAATAAAGTGTGCCACCACCAAGTTGGT	114
Dd	1586 TGCATAAGTAGCTGTCGAATTTCCACAAGCTTAGAATATAAATATCTTCCACCGGTGTA	1645
Oy	115 AACCTTAGGGTTTAAGACAGTCAAGATC	143
Dd	1646 AACTTAGGGTTTAAATGATATAATATC	1674
 RESULT 4 ABO91793		
LOCUS	ABO91793	5712 bp DNA linear MAM 02-MAY-2003
DEFINITION	Equus caballus AMELX gene for amelogenin, partial cds.	
ACCESSION	ABO91793	
VERSION	ABO91793.1 GI:29126038	
KEYWORDS		
SOURCE	Equus caballus (horse)	
ORGANISM	Equus caballus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.	
REFERENCE	1 Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N. From the Cover: The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)	
AUTHORS	2 (bases 1 to 5712) Iwase,M., Satta,Y. and Takahata,N. Direct Submission Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokendai), Department of Biosystems Science, Shonan Kofunshimura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwase@mekoryu.v01.soken.ac.jp, Tel:81-468-58-1571, Fax:81-468-58-1544)	
JOURNAL	Location/Qualifiers	
FEATURES	1..5712	
source	/organism="Equus caballus" /mol_type="genomic DNA" /db_xref="taxon:9796" /sex="male" 174..5712 /gene="AMELX" join(174..233,1578..1589) /gene="AMELX" join(1590..1643,3566..3613,4993..5037,5305..>5712) /gene="AMELX" /codon_start=1 /product="amelogenin" /protein_id="BA66112.1" /db_xref="gi:29126038" /translation="MGTWILPASLIIGAAFAIPDPSHGPEYINFSEVILTPLKWTQS LIRQPTSYGYEGPMGALHNIQIVLSIQONSNNALQPHNHI PMVSQHDPVPQQPMM PLPGCHSWPTQHHQPNLPBPVOQPFHQFQVQPQHPRIOPRPIDPIQLPQPRP PIPLQPRPLPRLPLEAMPATEQ"	
CDS	5'UTR	
gene		
ORIGIN		
Query Match	58.3%; Score 83.4; DB 4; Length 5712;	
Best Local Similarity	79.7%; Pred. No.1.8e-15;	
Matches 114; Conservative	0; Mismatches 21; Indels .8; Gaps 1;	
Oy	1 GATTTGTGGCGCCCTCTGGAGGACGCTTTGCTATGCGCGGTGAATAAACCCCT	60
Dd	1601 GATTTTGTGGCGACGCTCTGGAGGACGCTTTGCTATACCCGTGATATGATC----	1656
Oy	61 GCATTAATTTCAATTTCAAGAAGTTGGAATAAAGTGTGCCACCACCAAGTTGTAATCTTT	120
Dd	1657 ----AAATCCAAATTTCCAAAGCTTTGGAATAATCAAATCTGCTCCAAGCTTTGATAAATCTTT	1712
Oy	121 AGGGTTTAAGACAGTCAAGATC	143

Db	1713	AGCGTTAAGACAGTACAAGATC	1735
RESULT 5	AB091781	6465 bp	DNA
LOCUS	AB091781		linear
DEFINITION	Pan troglodytes AMELX gene for amelogenin, partial cds.		PRI 02-MAY-2003
ACCESSION	AB091781		
VERSION	AB091781.1	GI:29126015	
KEYWORDS			
SOURCE			
ORGANISM	Pan troglodytes (chimpanzee)		
REFERENCE	Pan troglodytes		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
TITLE	1		
JOURNAL	Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.		
MEDLINE	From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species		
PUBMED	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)		
REFERENCE	22608569		
AUTHORS	12672962		
TITLE	2 (bases 1 to 6465)		
JOURNAL	Iwase, M., Satta, Y. and Takahata, N.		
DEFINITION	Direct Submission		
ACCESSION	Submitted (19-SEP-2002)		
VERSION	Mineyo Iwase, Graduate University for		
KEYWORDS	Advanced Studies (Sokenai), Department of Biosystems Science;		
SOURCE	Shonan kokuusaimura, Hayama, Kanagawa 240-0193, Japan		
ORGANISM	(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel:81-468-58-1571,		
REFERENCE	Fax:81-468-58-1544)		
AUTHORS			
TITLE	Location/Qualifiers		
JOURNAL	1. 6465		
DEFINITION	/organism="Pan troglodytes"		
ACCESSION	/mol_type="genomic DNA"		
VERSION	/db_xref="taxon:9598"		
KEYWORDS	/sex="male"		
SOURCE	927. 6465		
REFERENCE	/gene="AMELX"		
AUTHORS	join(927. 982,2283. 2294)		
TITLE	/gene="AMELX"		
JOURNAL	join(2295. 2248,4284. 4331,5607. 5648,5740. 55781,		
DEFINITION	6052. 56465)		
ACCESSION	/gene="AMELX"		
VERSION	/codon_start=1		
KEYWORDS	/product="amelogenin"		
SOURCE	/protein_id="BAC66101.1"		
REFERENCE	/db_xref="GI:29126016"		
AUTHORS	/translation="WGTVILPACILGAAPMPLEPHRPGHYINFSYENSHQAINVD		
TITLE	RTVALPLAKMYOSIRPPYPSYGERPMGLHNOIIPVLSQQRPTTLOPHHII PVV		
JOURNAL	PAQCPVLPQOPMMPVPCQHSWTPDIOHQPLPFAQCPQOPVOPQOPHOPMOPQPV		
DEFINITION	HPMQRLPQOPPLPMPFMPQPLPMLPDLTEAMPSTDXK"		
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			

```

DEFINITION Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
ACCESSION AY040206
VERSION AY040206.1 GI:15028582
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 8810)
AUTHORS Hart, P.S., Hart, T.C., Simmer, J.P. and Wright, J.T.
TITLE A nomenclature for X-linked amelogenesis imperfecta
JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)
MEDLINE 21920287
PUBMED 11922868
REFERENCE 2 (bases 1 to 8810)
AUTHORS Hart, S., Hart, T.C. and Wright, T.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA
FEATURES
source location/Qualifiers
1..8810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.3-p22.1"
1..8791
/gene="AMELX"
1..1416
/promoter
/gene="AMELX"
1386..11391
/gene="AMELX"
join(1417..1472,2781..2846,4783..4830,6240..6281,
6552..6977,8611..8791)
/gene="AMELX"
/product="amelogenin precursor"
1417..1472
/gene="AMELX"
/number=1
2781..2846
/gene="AMELX"
/number=2
join(2793..2846,4783..4830,6240..6281,6552..6977,
8611..8616)
/gene="AMELX"
/codon_start=1
/product="amelogenin precursor"
/db_xref="GI:15028583"
/translation="MGTWILFACILGAAPFPPHGHGVINFSYEVLTPLKWTOS
IRPPYSYGVEMGWLHNOIIVLSQHPHTTLQPHHPIVPAQOQVTPQPMMP
VPGQSMTPLOHONPLPPAQQPQPOPOPOPHQPMQQRVNHMQPLPPQPLPP
MPPMQPLPMLPLPLTEANPSTDKREYD"
2793..2840
/gene="AMELX"
4783..4830
/gene="AMELX"
/number=3
6107..6148
/gene="AMELX"
/number=4
6240..6281
/gene="AMELX"
/number=5
6552..6977
/gene="AMELX"
/number=6
8611..8791
/gene="AMELX"
/number=7
8785..8791
/gene="AMELX"
sig_peptide
exon
exon
exon
exon
exon
exon
polya_signal
ORIGIN

```

```

Query Match 52.2% Score 74.6; DB 9; Length 8810;
Best Local Similarity 78.5%; Pred. No. 1.1e-12;
Matches 117; Conservative 0; Mismatches 24; Indels 8; Gaps 2;

QY 1 GATTTGTTTGGCTGCTCTGCGAGCAGCCCTTGTAGCCCGAGTAAATACCCCT 60
DB 2804 GATTTATTTGCTGCTCTGCGAGCAGCCCTTGTAGCCCGAGTAAATACCCCT 2863
QY 61 -----GCATTAATATTCATTTACAGAGCTGGAATAAAGCTGCCACAGTTGGT 114
DB 2864 TGCAATAGCAGATGTCATTTACAAACTTGACATTAATCTGCTC--ATAGTTGCT 2921
QY 115 AACTTAGGCTTTAAGACAGTAAAGATC 143
DB 2922 GAAATTAGGCTTTAAGACAGTAAAGATC 2950

RESULT 7
AL357130/c
LOCUS AL357130 158142 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.
ACCESSION AL357130
VERSION AL357130.3 GI:9863814
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1
REFERENCE Direct Submission
AUTHORS McIay, K.
TITLE Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9214076.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: b169E15
----- Summary Statistics
Assembly program: XGAP4; Version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 156466 bases at least Q30
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality
coverage: 4.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68897: contig of 68897 bp in length
* 68898 68997: gap of 100 bp
* 68998 122842: contig of 53845 bp in length
* 122843 122942: gap of 100 bp
* 122943 123584: contig of 6642 bp in length
* 123585 123684: gap of 100 bp
* 123685 158142: contig of 28458 bp in length.
FEATURES
source location/Qualifiers
1..158142
/organism="Homo sapiens"

```



```

repeat_region      /rpt_family="Alu"
                    complement(30217..30408)
repeat_region      /rpt_family="L1M4S"
                    /rpt_family="Alu"
                    complement(31199..31364)
repeat_region      /rpt_family="Alu"
                    complement(33939..34460)
repeat_region      /rpt_family="L1MA2"
                    complement(34952..35384)
repeat_region      /rpt_family="MLT1d"
                    complement(36035..36126)
repeat_region      /rpt_family="MIR2"
                    complement(36850..37035)
repeat_region      /rpt_family="MER20"
                    complement(38042..38250)
STS
repeat_region      /db_xref="dbSTS:L47508"
                    complement(38122..38510)
repeat_region      /rpt_family="MSTa"
                    complement(38797..40103)
repeat_region      /rpt_family="MIR"
                    complement(42465..42539)
repeat_region      /rpt_family="MIR"
                    complement(47139..47335)
repeat_region      /rpt_family="MER3"
                    complement(48993..49166)
repeat_region      /rpt_family="MIR"
                    complement(49718..49918)
repeat_region      /rpt_family="L1PAl1"
                    complement(51943..52505)
repeat_region      /rpt_family="L1MB3"
                    complement(53226..53677)
repeat_region      /rpt_family="MLT1d"
                    complement(54846..54966)
repeat_region      /rpt_family="MIR"
                    complement(55040..55226)
repeat_region      /rpt_family="MER20"
                    complement(55691..55837)
repeat_region      /rpt_family="Alu"
                    complement(56051..56486)
repeat_region      /rpt_family="L1MB3a"
                    complement(58557..59030)
repeat_region      /rpt_family="MLT1d"
                    complement(59270..59428)
repeat_region      /rpt_family="MERS"
                    complement(59847..60030)
repeat_region      /rpt_family="MER20"
                    complement(61290..61384)
repeat_region      /rpt_family="L1B1"
                    complement(61883..61811)
repeat_region      /rpt_family="MER20"
                    complement(61907..62113)
repeat_region      /rpt_family="L1MB7"
                    complement(63588..63969)
repeat_region      /rpt_family="L1MA10"
                    complement(64800..65073)
repeat_region      /rpt_family="Alu"
                    complement(67157..67209)
repeat_region      /rpt_family="MIR"
                    complement(67637..75845)
gene               /gene="HAMELX"
                    complement(67637..75845)
promoter           /gene="HAMELX"
                    complement(67637..75845)
                    /note="similar to bovine amelogenin promoter"
                    join(68495..68552,69861..69926,71863..71910,73320..73361,
                    73632..74057,75691..75845)
                    /gene="HAMELX"
                    /product="amelogenin"
                    join(69873..69826,71863..71910,73320..73361,73632..74057,
                    75691..75696)
                    /gene="HAMELX"
                    /note="similar to amelogenin encoded by GenBank Accession
                    Number M86932"
                    /codon_start=1
                    /product="amelogenin"

```

```

/protein_id="AAC21581.1"
/db_xref="GI:2739350"
/translation="MGTWILPACLLGAAPMLPPHGHGXYINSYEVLPPLKMYQS
IRPPYSGYEPMGMLHNOIIPVLSOCHPPTHTLQPHNHIPVVPAAQPVLEQPMMP
VGGOSMTPIQHNOPNLPRPAQOPVQOPVOPQPHQMPQPPVHMPQLPRPPLRP
MPPQPLPMLPDLTLEAMPSTDKKREBV"
complement(77161..77331)
/rpt_family="MER3"
complement(79184..79473)
/rpt_family="Alu"
complement(79881..80169)
/rpt_family="Alu"
complement(80253)
/rpt_family="MER20"
complement(80384..80459)
/rpt_family="L1PB3"
complement(81544..81633)
/rpt_family="MIR"
complement(84789..84957)
/rpt_family="L1MA2"
complement(84978..85252)
/rpt_family="Alu"
complement(85294..85692)

```

Query Match 52.2% Score 74.6; DB 9; Length 259202;
 Best Local Similarity 78.5%; Pred. No. 1,1e-12;
 Matches 117; Conservative 0; Mismatches 24; Indels 8; Gaps 2;

```

Qy 1 GATTTGGTTGGCTGCTCTCGGAGCAGACCTTTGCTATGCCGTAATAATCCCT 60
Db 69884 GATTTATTTGGCTGCTCTCGGAGCAGACCTTTGCTATGCCGTAATAATCCCT 69943
Qy 61 -----GCATATATTCATTTCAACACTTGGAATAAGTCCGCCACCAATTGGT 114
Db 69944 TGCATTAAGTCAGTGTCCATTTCAACACTTGGAATAATCTGCTC--ATACTGGT 70001
Qy 115 AACTTAGGGTTTAAGACATCAAGATC 143
Db 70002 GAAATTAGGCTTTAAACAGTATGAGATC 70030

```

RESULT 9
 BV089295/c 695 bp DNA linear STS 15-OCT-2003
 LOCUS RPAWMSR00001191 Roche Palo Alto Mus musculus STS genomic, sequence
 DEFINITION tagged site.
 ACCESSION BV089295
 VERSION BV089295.1 GI:37666774
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 695)
 Usuka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
 McPherson,J.D., Fournzler,D. and Peltz,G.
 Mus musculus SNPs
 Unpublished (2003)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Jonathan Usuka
 Roche Palo Alto Genetics and Genomics Department
 Roche Palo Alto
 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
 Tel: 6508555807
 Email: Jonathan.Usuka@roche.com
 Primer A: No primer submitted
 Primer B: No primer submitted.
 location/Qualifiers

FEATURES
 source
 1..695
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="X-73613-72919-AL805974.7.1.95826"

OY 117 CTTAGGTTTAAAGACAGTACAGAT 142
 |||||
 Db 2440 AGTTAGGTTTAAACGTTATGAT 2465

RESULT 12
 D83063S2 463 bp DNA linear ROD 26-NOV-2003
 LOCUS Mus musculus DNA for amelogenin, exon 2.
 DEFINITION D83064
 VERSION D83064.1 GI:2687869
 KEYWORDS 2 of 5
 SEGMENT Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Oida,S., Miyazaki,H., Iimura,T., Suzuki,M., Sasaki,S. and Shimokawa,H.
 TITLE Molecular structure of the mouse amelogenin genomic DNA
 JOURNAL DNA Seq. 6 (5), 307-310 (1996)
 MEDLINE 97142134
 PUBMED 8988368
 REFERENCE 2 (bases 1 to 463)
 AUTHORS Oida,S.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1996) Shinichi Oida, Tokyo Medical and Dental University, Fac. Dentistry, Dept. Biochemistry, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan (E-mail:s.oida.obchdent.tmd.ac.jp, Tel:03-5803-5448, Fax:03-5803-0187)
 COMMENT On Dec 15, 1997 this sequence version replaced gi:2662357.

FEATURES

source
 1..463
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /sub_species="Balb-C"
 /db_xref="taxon:10090"
 /chromosome="X"
 /map="Xp3-F4"
 /clone="MAMEG-2"
 /cell_type="Leukocyte"
 /clone_lib="EMBL-3 SP6-T7"
 /dev_stage="adult"
 join(D83063.1:203..258,222..233)
 <1..221
 /number=1
 222..287
 /product="amelogenin"
 /number=2
 288..463
 /number=2
 Intron
 5'UTR
 Intron
 exon
 Intron

ORIGIN

Query Match 50.5%; Score 72.2; DB 10; Length 463;
 Best Local Similarity 77.2%; Pred. No. 6,1e-12;
 Matches 115; Conservative 0; Mismatches 28; Indels 6; Gaps 2;

OY 1 GATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCGTAGTAAATACCCC-59
 |||||
 Db 245 GATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCGTAGTAAATACCCA 304
 |||||
 OY 60 ----TGCATTAATTCATTTTCAAGCTTGAATAAAGTGTGCCCAACGATTGTA 115
 |||||
 Db 305 TTACTAATTCATTCATTTTCAAGCTTGAATAAAGTGTGCCCAACGATTGTA 364
 |||||
 OY 116 ACTTT-AGGTTTAAAGACAGTACAGATC 143
 |||||
 Db 365 ATTTCACTGTGTAAACAGTACAGATC 393
 |||||

RESULT 13
 AF294397

LOCUS AF294397 9384 bp DNA linear ROD 03-SEP-2000
 DEFINITION Mus musculus amelogenin gene, promoter and partial cds.
 ACCESSION AF294397
 VERSION AF294397.1 GI:9965405
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Sneed,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P., Paine,C.T., Chen,L.S., Burslein,J.M., Jitpukdeeputintara,S., White,S.N. and Bringsas,P. Jr.
 TITLE Transgene animal model for protein expression and accumulation into forming enamel
 JOURNAL Connect. Tissue Res. 38 (1-4), 279-286 (1998)
 MEDLINE 20515040
 PUBMED 11063035

REFERENCE
 AUTHORS Sneed,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
 TITLE Direct Submission
 JOURNAL Submitted (08-AUG-2000) Dentistry, University of Southern California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
 COMMENT Location/Qualifiers
 1..9384
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 1..8546
 7188..7194
 8547..8612
 /product="amelogenin"
 8559..8612
 /note="enamel matrix protein"
 /codon_start=1
 /product="amelogenin"
 /protein_id="AAG10078.1"
 /db_xref="GI:9965405"
 /translation="MGWILFACLLGAAMPMP"

Promoter
 TATA_signal
 mRNA

CDS
 8559..8612
 /product="amelogenin"

ORIGIN

Query Match 50.5%; Score 72.2; DB 10; Length 9384;
 Best Local Similarity 77.2%; Pred. No. 6,3e-12;
 Matches 115; Conservative 0; Mismatches 28; Indels 6; Gaps 2;
 OY 1 GATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCGTAGTAAATACCCC-59
 |||||
 Db 8570 GATTTGTTGCTGCTCTCGGAGACGCTTGTATGCCGTAGTAAATACCCA 8629
 |||||
 OY 60 ----TGCATTAATTCATTTTCAAGCTTGAATAAAGTGTGCCCAACGATTGTA 115
 |||||
 Db 8630 TTACTAATTCATTCATTTTCAAGCTTGAATAAAGTGTGCCCAACGATTGTA 8689
 |||||
 OY 116 ACTTT-AGGTTTAAAGACAGTACAGATC 143
 |||||
 Db 8690 ATTTCACTGTGTAAACAGTACAGATC 8718
 |||||

RESULT 14
 AL805974/c

LOCUS AL805974 95826 bp DNA linear ROD 29-AUG-2003
 DEFINITION Mouse DNA sequence from clone RP23-334F21 on chromosome X, complete sequence.
 ACCESSION AL805974
 VERSION AL805974.8 GI:34366495
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Chapman,J.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 30, 2003 this sequence version replaced g1:25955748.
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Mp, MORNPEP; Information on the MORNPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-334F21 is from the RP23 Mouse BAC Library constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBac3.6.

FEATURES

Source

Location/Qualifiers
1..95826

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="x"
/clone="RP23-334F21"
/clone_id="RP23-334F21"

ORIGIN

Query Match 50.5%; Score 72.2; DB 10; Length 95826;
Best Local Similarity 77.2%; Pred. No. 6.4e-12;
Matches 115; Conservative 0; Mismatches 28; Indels 6; Gaps 2;
QY 1 GATTTGTTGCTGCTCTGAGAGAGAGCTTTGCTATGCCCGTGAATAATACCCC-59
DB 22692 GATTTGTTGCTGCTCTGAGAGAGAGCTTTGCTATGCCCGTGAATAATACCA 22633
QY 60 ----TGCAATATATTCATTTACACAGCTTGAATAAAGTCTGCCCAACGATTGTA 115
DB 22632 TTACTATATTCATTCATTTACACAGCTTGAATAAAGTCTGCCCAACGATTGTA 22573
QY 116 ACTTT-AGGTTTAAGACAGTACAGATC 143
DB 22572 ATTTCCTGTGTACACAGTACAGATC 22544

RESULT 15
AB091791 7425 bp DNA linear MAM 02-MAY-2003
LOCUS AB091791
DEFINITION Sus scrofa AMELX gene for amelogenin, partial cds.
ACCESSION AB091791 GI:29126034
VERSION AB091791.1
KEYWORDS
SOURCE Sus scrofa (pig)

ORGANISM

REFERENCE

AUTHORS

TITLE

1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

JOURNAL

MEDLINE

PUBMED

12672962

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@shonan.kokusaiinura.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES

Source

Location/Qualifiers
1..7425

/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/sex="male"
1491..7425
/gene="AMELX"
join(1491..1548,3140..3151)
/gene="AMELX"
join(3152..3205,5316..5363,6619..6660,6717..6761,
7033..7425)
/gene="AMELX"

gene

5' UTR

CDS

ORIGIN

Query Match 49.7%; Score 71; DB 4; Length 7425;
Best Local Similarity 72.5%; Pred. No. 1.5e-11;
Matches 108; Conservative 0; Mismatches 35; Indels 6; Gaps 1;
QY 1 GATTTGTTGCTGCTCTGAGAGAGAGCTTTGCTATGCCCGTGAATAATACCCC 60
DB 3163 GATTTGTTGCTGCTCTGAGAGAGAGCTTTGCTATGCCCGTGAATAATACCA 3222
QY 61 GCATA-----ATATTCATTTACACAGCTTGAATAAAGTCTGCCCAACGATTGTA 114
DB 3223 TTGTAAGTACAGTCAATGATCATAGCTTGAATAAATCTACCCCGAGTTGATA 3282
QY 115 AACTTAGGTTTAAGACAGTACAGATC 143
DB 3283 CACGTTAGGATGAAAACAGTACAGATC 3311

Search completed: February 8, 2005, 16:37:00
Job time: 798.321 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 : Search time 189.696 Seconds
(without alignment)
3957.220 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143
Sequence: 1 gattctgttcgtcctgcctcc.....gtttaagacagtaacagatc 143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_GeneSeq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.6	52.2	259202	12	ADQ18492
2	40.4	28.3	270	12	ADM80835
3	40.4	28.3	549	12	ADM80826
4	40.4	28.3	623	12	ADM80827
5	40.2	28.1	65	6	ABN28389
6	40.2	28.1	476	3	AAZ50832
7	39.2	27.4	556	10	ADB59026
8	39.2	27.4	556	10	ADB53782
9	38.8	27.1	722	6	AA141111
10	38.8	27.1	752	6	AA141110
11	37.2	26.0	318	12	ADM80836
12	37.2	26.0	750	2	AAZ07020
13	37.2	26.0	793	2	AAZ07018
14	35.6	24.9	802	2	AAZ07019
15	35.6	24.9	852	12	ADQ22977
16	30	21.0	1284	9	ADA30717
17	30	21.0	47804	10	ADC86176
18	30	21.0	110000	2	AA151990_09
19	29.8	20.8	237	4	AA121507
20	29.8	20.8	237	4	ABA65686
21	29.8	20.8	237	4	AA146800

22	29.8	20.8	237	4	ABA48672	ABA48672 Human bre
23	29.8	20.8	237	4	ABA33654	ABA33654 Probe #12
24	29.8	20.8	237	4	AAK40744	AAK40744 Human bon
25	29.8	20.8	237	4	AAK15013	AAK15013 Human bra
26	29.8	20.8	237	4	ABS40318	ABS40318 Human liv
27	29.8	20.8	237	5	AA107203	AA107203 Probe #71
28	29.8	20.8	237	6	ABS14698	ABS14698 Human gen
29	29.8	20.8	520	6	ABO56275	ABO56275 Human ova
30	29.8	20.8	843	5	AA868794	AA868794 DNA encod
31	29.8	20.8	1513	6	AB216622	AB216622 Arabidops
32	29.8	20.8	1513	8	ADA69305	ADA69305 Arabidops
33	29.8	20.8	60327	9	ADA02681	ADA02681 Mouse dus
34	29.8	20.8	60327	10	ADB72419	ADB72419 Mouse dus
35	29.8	20.8	60327	10	ADB59292	ADB59292 Mouse dus
36	29.8	20.7	116592	8	ABX15519	ABX15519 Soybean t
37	29.6	20.7	116592	10	AAAD47900	AAAD47900 Human tyr
38	29.6	20.6	201	4	AA125916	AA125916 Human tra
39	29.4	20.6	417	4	AA125839	AA125839 Human bre
40	29.4	20.6	430	4	AA119004	AA119004 Human bre
41	29.4	20.6	434	4	AA118062	AA118062 Human bre
42	29.4	20.6	446	4	AA108094	AA108094 Human bre
43	29.4	20.6	558	4	AA119081	AA119081 Human bre
44	29.4	20.6	676	10	AB283542	AB283542 Toxicolog
45	29.4	20.6				

ALIGNMENTS

RESULT 1	ADQ18492	standard; DNA; 259202 BP.
ID	ADQ18492	
XX	ADQ18492	
AC	ADQ18492	
XX	ADQ18492	
DT	26-AUG-2004	(first entry)
XX	26-AUG-2004	
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KW	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
XX	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
OS	Homo sapiens.	
XX	Homo sapiens.	
XX	10-JUN-2004.	
PD	10-JUN-2004.	
XX	10-JUN-2004.	
PF	26-NOV-2003; 2003MO-US038193.	
XX	26-NOV-2003; 2003MO-US038193.	
PR	26-NOV-2002; 2002US-0429739P.	
XX	26-NOV-2002; 2002US-0429739P.	
PA	(PROT-) PROTEIN DESIGN LABS INC.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PI	Aziz N, Gineburg WM, Zlotnik A;	
XX	Aziz N, Gineburg WM, Zlotnik A;	
DR	WPI; 2004-441208/41.	
XX	WPI; 2004-441208/41.	
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210pp; English.	
PS	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	
CC	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	
CC	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	
CC	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

PT	connective tissue and cell proliferative disorders including cancer.
PS	Claim 5; SEQ ID NO 55; 272pp; English.
XX	
CC	The present sequence encodes a human cell adhesion and extracellular
CC	matrix protein designated CADECM. CADECM sequences has neuroprotective,
CC	cytostatic and anorectic activites. The CADECM polypeptides and
CC	polynucleotides are useful in diagnosing, treating and preventing immune
CC	neurological, developmental, connective tissue and cell proliferative
CC	disorders including cancer, e.g. breast, prostate, ovarian, lung or colorectal
CC	cancer, obesity and Tangier disease.
XX	
SO	Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
Qy	Query Match 28.3%; Score 40.4; DB 12; Length 549;
Be	Best Local Similarity 88.0%; Pred. No. 0.0012;
Ma	Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0
Ds	1 GATTTCGTTGCTGCTCCTCGGACAGCCTTGCATGCCGTAGTA 50
	47 GATTTCATTTGCTGCTCCTCGGACAGCCTTGCATGCCGTGCTTA 96
XX	
RESULT 4	
ADM80827	
ID	ADM80827 standard; cDNA; 623 BP.
AC	ADM80827;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human CADECM-14 encoding cDNA SEQ ID NO:56.
XX	
KW	human; cell adhesion and extracellular matrix protein; CADECM;
KW	neuroprotective; cytostatic; anorectic; immune disorder;
KW	neurological disorder; developmental disorder;
KW	connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX	Tangier disease; gene; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key Location/Qualifiers
FT	CDS 36..563
FT	/tag= a
FT	/product= "CADECM-14"
XX	
PN	MO2004015396-A2.
XX	
PD	19-FEB-2004.
XX	
PF	12-AUG-2003; 2003WO-US025418.
XX	
PR	13-AUG-2002; 2002US-0403781P.
PR	30-AUG-2002; 2002US-0407034P.
PR	13-SEP-2002; 2002US-0410566P.
PR	24-SEP-2002; 2002US-0413482P.
PR	25-SEP-2002; 2002US-0413890P.
PR	08-NOV-2002; 2002US-0424904P.
PR	13-NOV-2002; 2002US-0426222P.
XX	
PA	(INCY-) INCYTE CORP.
XX	
PI	Elloct VS, Khare R, Emerling BM, Kable AB, Tran UK, Jin P;
PI	Becia SD, Margule JP, Swarnakar A, Chewla NK, Ramkumar J;
PI	Hafalla AD, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
PI	Wang JT, Chien D, Yang YG;
XX	
DR	WPI; 2004-191795/18.
XX	
DR	P-PsDB; ADM80785.
XX	
PT	New cell adhesion and extracellular matrix proteins, useful in
PT	diagnosing, treating and preventing immune, neurological, developmental,
PT	connective tissue and cell proliferative disorders including cancer.

Claim 5; SEQ ID NO 56; 272bp; English.
 The present sequence encodes a human cell adhesion and extracellular matrix protein designated CADCEM. CADCEM sequences has neuroprotective, cytoskeletal and anorectic activities. The CADCEM polypeptides and polynucleotides are useful in diagnosing, treating and preventing immune, neurological, developmental, connective tissue and cell proliferative disorders including cancer, e.g. breast, prostate, ovarian, lung or colon cancer, obesity and Tangier disease.
 Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
 Query Match 28.3%; Score 40.4; DB 12; Length 623;
 Best Local Similarity 88.0%; Pred. No. 0.0013;
 Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0
 1 GATTTGTTGGCTGCTGCTGCGAGGACGACCTTGCTATGCGCGTAGTA 50
 47 GATTTTATTGGCTGCTGCTGCTGCGAGACACTTTGGCATGCTGTGCTTA 96
 RESULT 5
 ABN28389
 ID ABN28389 standard; DNA; 65 BP.
 AC ABN28389;
 DT 15-JUL-2002 (first entry)
 DE Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.
 OS Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX Rattus norvegicus.
 WO200210449-A2.
 07-FEB-2002.
 20-JUL-2001; 2001WO-1B001903.
 28-JUL-2000; 2000US-0221607P.
 02-MAY-2001; 2001US-0287724P.
 (COMP-) COMPUGEN INC.
 Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 WPI; 2002-257383/30.
 New oligonucleotide libraries comprising oligonucleotides which selectively hybridize to mRNAs transcribed from a transcription unit of a genome, useful for detecting tissue-, pathology-, and developmental-specific genes.
 Example 1; SEQ ID NO 1137; 47bp; English.
 The present invention describes oligonucleotide libraries for detecting messenger RNAs that populate a (sub-)transcriptome, where the (sub-)transcriptome comprises messenger RNAs transcribed from multiple transcription units that populate a genome. The library comprises several oligonucleotides, each capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, which encodes one or more messenger RNA splice variants. The oligonucleotide libraries are useful for detecting mRNAs from a biological sample, in expression profiling studies, in qualitatively or quantitatively characterizing the corresponding transcriptome, and in detecting RNA transcripts and splice variants of human or animal transcriptomes. The libraries may also be used as specialised mini libraries to detect transcripts of a sub-transcriptome under a particular biological or pathological state, and so allowing the detection of tissue

CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pcc_sequences

XX
SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;

Query Match 28.1%; Score 40.2; DB 6; Length 65;
Best Local Similarity 93.3%; Pred. No. 0.00067;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTGAGAGACGCTTTGCTATGCCCGT 45
6 GATCTTGTGGCTGCTCTGAGAGACGCTTTGCTATGCCCGT 50

Db

RESULT 6
AAZ50832
ID AAZ50832 standard; DNA; 476 BP.
XX
XX AAZ50832;
XX
DT 31-MAY-2000 (first entry)
XX
XX Rat amelogenin gene (A4).
XX
XX Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;
XX chondrogenic inducing molecule; CIM; cartilage growth; osteopthic;
XX extracellular matrix protein; tooth enamel; enamel mineralisation;
XX ameloblast; bone regeneration; composite cell construct; ds.
XX
XX Ratus sp.
OS

XX
FH Key Location/Qualifiers
FT exon 1..36
FT /tag= a
FT /partial
FT /number= 1
FT 37..101
FT /tag= b
FT /number= 2
FT 48..317
FT /tag= c
FT /product= "Rat amelogenin protein"
FT /transl_except= (pos:69..71, aa:Gly)
FT 48..95
FT /tag= d
FT 96..314
FT /tag= e
FT /label= Mature_rat_amelogenin_protein
FT 102..149
FT /tag= f
FT /number= 3
FT 150..191
FT /tag= g
FT /number= 4
FT 192..236
FT /tag= h
FT 237..311
FT /tag= i
FT /number= 6d
FT /note= "Comprises of gene segments 6a, b, c and d"
FT 312..317
FT /tag= j
FT /number= 7
FT /note= "includes the stop codon"
XX
XX

PN WO200006734-A1.
XX
XX 10-FEB-2000.
XX
XX 29-JUL-1999; 99WO-US017342.
XX
XX 29-JUL-1998; 98US-0094489P.
XX
XX (NOUN) UNIV NORTHWESTERN.
XX
XX Vels A, Nebgen DR;
XX
XX WPI: 2000-205464/18.
XX P-PSDB; AA45074.
XX
XX Novel amelogenin polypeptides and polynucleotides, useful for enhancing
XX bone generation in mammals and synthesizing bone matrix or articular
XX surfaces at implant sites.
XX
XX Example 2; Fig 11B; 79pp; English.
XX
XX The present DNA sequence is the full-length rat amelogenin gene (A4),
XX comprising exons 1-7, including the exon segment 6d. It is derived from
XX the rat incisor odontoblast-pulp cDNA library. The splice variants of
XX this gene functions as an osteogenic or chondrogenic inducing molecule
XX (CIM), which is useful for enhancing bone or cartilage growth. It has
XX osteopathic activity. Amelogenin belongs to the family of extracellular
XX matrix proteins, in developing tooth enamel, that are produced by the
XX CC osteogenic inducing amelogenin molecules are useful to induce
XX CC differentiation of cells to the osteogenic and chondrogenic phenotypes
XX and can be used in a composite cell construct for bone and cartilage
XX regeneration. The polynucleotides can be employed to produce the
XX polypeptides by recombinant techniques

SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;

Query Match 28.1%; Score 40.2; DB 3; Length 476;
Best Local Similarity 93.3%; Pred. No. 0.0014;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTGAGAGACGCTTTGCTATGCCCGT 45
59 GATCTTGTGGCTGCTCTGAGAGACGCTTTGCTATGCCCGT 103

Db

RESULT 7
ADB59026
ID ADB59026 standard; DNA; 556 BP.
XX
XX ADB59026;
XX
DT 04-DEC-2003 (first entry)
XX
XX Toxicity-related gene, SEQ ID 4052.
XX
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
XX drug screening; toxicity assay; ds.
XX
XX Unidentified.
XX
XX WO2003064624-A2.
XX
XX 07-AUG-2003.
XX
XX 31-JAN-2003; 2003WO-US003194.
XX
XX 31-JAN-2002; 2002US-00060087.
XX 15-MAR-2002; 2002US-0364045P.
XX 15-MAR-2002; 2002US-0364055P.
XX 30-DEC-2002; 2002US-0436643P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX

XX	PI	Mendrick D, Porter M, Johnson K, Higge B, Castle A, Elashoff M,	
XX	DR	WPI; 2003-689530/65.	
XX	XX		
PT	PT	Predicting a toxic effect of a compound, useful in identifying toxicity	
PT	PT	markers in liver tissues or cells for drug screening and toxicity assays,	
PT	PT	completes preparing gene expression profile of tissue or cells exposed to	
XX	XX	the compound.	
PS	PS	Claim 1; SEQ ID NO 4052; 1156bp; English.	
XX	XX		
CC	CC	The present invention relates to a method for predicting a toxic effect	
CC	CC	of a compound. The method comprises preparing a gene expression profile	
CC	CC	of a tissue or cell sample exposed to the compound, and comparing the	
CC	CC	gene expression profile to a database comprising SEQ ID 1-4925, where	
CC	CC	differential expression of the gene indicates at least one toxic effect.	
CC	CC	The method is useful for predicting at least one toxic effect of a	
CC	CC	compound, predicting hepatotoxicity or the progression of a toxic effect	
CC	CC	of a compound, identifying an agent that modulates the onset or	
CC	CC	progression of a toxic response, predicting the cellular pathways that a	
CC	CC	compound modulates in a cell, and identifying an agent that modulates at	
CC	CC	least one activity of a protein. The method and compositions of the	
CC	CC	present invention using a database of genes having liver toxin-induced	
CC	CC	differential expression, are useful in identifying toxicity markers in	
CC	CC	liver tissues or cells for drug screening and toxicity assays. Note: The	
CC	CC	sequence data for this patent did not form part of the printed	
CC	CC	specification, but was obtained in electronic format directly from WIPO	
XX	XX	at ftp.wipo.int/pub/published_pct_sequences.	
SO	SO	Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;	
		Query Match 27.4%; Score 39.2; DB 10; Length 556;	
		Best Local Similarity 91.1%; Pred. No. 0.0032;	
		Matches 41; Conservative 0; Mismatches 4; Indels 0; Gaps 0	
Oy	Oy	1 GATTTTGTTGGCTTGGCTTCTGAGAGACGCTTTGCTATGCCCCGT 45	
Db	Db	33 GATCTTGTTCCTCTCTCTCTGAGACGCTTTGATATGCCCCCT 77	
RESULT 8			
ADBS3782			
ID	ID	ADBS3782 standard; DNA; 556 BP.	
XX	XX		
AC	AC	ADBS3782;	
XX	XX		
DT	DT	04-DEC-2003 (first entry)	
XX	XX		
DE	DE	Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.	
XX	XX		
KW	KW	toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;	
KW	KW	toxicity marker; toxicity progression; drug screening;	
KW	KW	primary rat hepatocyte toxicity modelling; gene; de.	
XX	XX		
OS	OS	Rattus norvegicus.	
XX	XX		
PN	PN	WO2003065993-A2.	
XX	XX		
PD	PD	14-AUG-2003.	
XX	XX		
PF	PF	04-FEB-2003; 2003WO-US003482.	
XX	XX		
PR	PR	04-FEB-2002; 2002US-0353171P.	
PR	PR	13-MAR-2002; 2002US-0365348P.	
PR	PR	08-APR-2002; 2002US-0370248P.	
PR	PR	10-APR-2002; 2002US-0371134P.	
PR	PR	10-APR-2002; 2002US-0371135P.	
PR	PR	11-APR-2002; 2002US-0371150P.	
PR	PR	11-APR-2002; 2002US-0371411P.	
PR	PR	19-APR-2002; 2002US-0373601P.	
PR	PR	19-APR-2002; 2002US-0373602P.	
PR	PR	22-APR-2002; 2002US-0374139P.	

PR	08-MAY-2002;	2002US-03786370P.
PR	09-MAY-2002;	2002US-0378652P.
PR	09-MAY-2002;	2002US-0378653P.
PR	09-MAY-2002;	2002US-0378665P.
PR	09-JUL-2002;	2002US-0394230P.
PR	09-JUL-2002;	2002US-0394253P.
PR	04-SEP-2002;	2002US-0407688P.
PR	28-JAN-2003;	2003US-0442900P.
XX	(GENE-) GENE LOGIC INC.	
PA	Mendrick D,	Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI	Elshof M;	
PI	WPI; 2003-731472/69.	
DR		
XX	Determining if a compound induces a toxic effect on a tissue or cell, for	
PT	identifying hepatotoxic compounds, comprises comparing a gene expression	
PT	profile of a tissue or cell sample to a database of Tox mean and non-Tox	
PT	mean values.	
PS	Claim 44; SEQ ID NO 4324; 874pp; English.	
XX	The present invention describes a method for determining whether a	
CC	compound induces a toxic effect on a tissue or cell. The method comprises	
CC	preparing a gene expression profile of a tissue or cell sample exposed to	
CC	the compound, and comparing the gene expression profile to a database	
CC	comprising data or information on the Tox mean and non-Tox mean value.	
CC	The method is useful for predicting or identifying at least one toxic	
CC	effect, particularly hepatotoxicity, of a test or unknown compound. The	
CC	genes listed in the specification are useful as diagnostic or toxicity	
CC	markers for the prediction or identification of the physiological state	
CC	of tissue or predict the toxic effects of a compound, or to	
CC	identify or predict the toxic effects of a compound or an agent. These	
CC	may also be used as markers for monitoring toxicity progression or for	
CC	drug screening. The present sequence represents a primary rat hepatocyte	
CC	toxicity modelling related gene sequence from the present invention.	
XX		
XX	Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;	
SQ		
Query Match	27.4%; Score 39.2; DB 10; Length 556;	
Best Local Similarity	91.1%; Pred. No. 0.0032;	
Matches 41; Conservative	0; Mismatches 4; Indels 0; Gaps 0;	
OY	1 GATTTTGTTCCTGCTCGTGAGACGCTTTGTAATGCCGT 45	
DB	33 GATCTTGTTCCTGCTCGTGAGACGCTTTGTAATGCCCT 77	
RESULT 9		
AL41111		
ID	AL41111 standard; DNA; 722 BP.	
XX		
AC	AA41111;	
XX		
DT	16-OCT-2002 (first entry)	
XX		
DE	GAMU related Y-chromosome DNA sequence.	
XX		
KM	Goat embryo sexual identification technique; goat amelogenin gene; GAMU;	
KW	sex-specific; gene; ds; Y-chromosome.	
XX		
OS	Capra hircus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	35..658
FT		/tag= a
XX		/product= "Y-chromosome protein"
PX	TMA54013-A.	
PD	11-SEP-2001.	
XX		

PF 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX (CHEN/) CHEN C.
XX (JANG/) JANG J.
XX (WENG/) WENG T.
XX (JENG/) JENG D.
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX WPI; 2002-442016/47.
XX P-PSDB; AAO22534.
XX
XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
PT identification, comprises high sensitivity even using single white blood
PT cell or cleavage c.
XX
XX Disclosure; Page 28; 35pp; Chinese.
XX
XX The invention relates to a goat embryo sexual identification technique
CC with high efficiency, sensitivity and repeatability. This technique
CC involves separately cloning and sequencing the coding regions and the
CC introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
CC results indicate that there are sex-specific sequences in the fifth
CC intron of the gene. The major characteristics according to the present
CC invention include high sensitivity, applicable in sex identification even
CC only using a single white blood cell or a single cleavage cell of
CC blastula; high diagnostic efficiency, capable of identifying hundreds of
CC goat embryo in 3 hours; simple operation procedures without complicated
CC steps of DNA extraction and need no additional control group intron; and
CC can be applied on different species of goats. This polynucleotide
CC sequence represents a gAML related X-chromosome DNA sequence of the
CC invention
XX
XX Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;
SQ
Query Match 27.1%; Score 38.8; DB 6; Length 722;
Best Local Similarity 95.2%; Pred. No. 0.0048;
Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GATTTTGTTGCTGCTCTGAGAGAGCCTTGTATGCC 42
DB 46 GATTTTGTTGCTGCTCTGAGAGAGCCTTGTATGCC 87
RESULT 10
ID AAL41110 standard; DNA; 752 BP.
XX
XX AAL41110;
AC
XX
XX 16-OCT-2002 (first entry)
DT
XX
XX gAML related X-chromosome DNA sequence.
DE
XX
XX Goat embryo sexual identification technique; goat amelogenin gene; gAML;
KW sex-specific; gene; db; X-chromosome.
XX
XX Capra hircus.
OS
XX
XX Key Location/Qualifiers
FH 35..658
FT /*tag= a
FT /product= "X-chromosome protein"
XX
XX TW454013-A.
XX
XX 11-SEP-2001.
PD
XX
XX 10-NOV-1999; 99TW-00119616.
PF
XX
XX 10-NOV-1999; 99TW-00119616.
PR

XX
XX (CHEN/) CHEN C.
XX (JANG/) JANG J.
XX (WENG/) WENG T.
XX (JENG/) JENG D.
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX WPI; 2002-442016/47.
XX P-PSDB; AAO22534.
XX
XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
PT identification, comprises high sensitivity even using single white blood
PT cell or cleavage c.
XX
XX Disclosure; Page 28; 35pp; Chinese.
XX
XX The invention relates to a goat embryo sexual identification technique
CC with high efficiency, sensitivity and repeatability. This technique
CC involves separately cloning and sequencing the coding regions and the
CC introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
CC results indicate that there are sex-specific sequences in the fifth
CC intron of the gene. The major characteristics according to the present
CC invention include high sensitivity, applicable in sex identification even
CC only using a single white blood cell or a single cleavage cell of
CC blastula; high diagnostic efficiency, capable of identifying hundreds of
CC goat embryo in 3 hours; simple operation procedures without complicated
CC steps of DNA extraction and need no additional control group intron; and
CC can be applied on different species of goats. This polynucleotide
CC sequence represents a gAML related X-chromosome DNA sequence of the
CC invention
XX
XX Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
SQ
Query Match 27.1%; Score 38.8; DB 6; Length 752;
Best Local Similarity 95.2%; Pred. No. 0.0049;
Matches 40; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GATTTTGTTGCTGCTCTGAGAGAGCCTTGTATGCC 42
DB 46 GATTTTGTTGCTGCTCTGAGAGAGCCTTGTATGCC 87
RESULT 11
ID ADM80836 standard; cDNA; 318 BP.
XX
XX ADM80836;
AC
XX
XX 03-JUN-2004 (first entry)
DT
XX
XX Human CADEC23 encoding cDNA SEQ ID NO:65.
DE
XX
XX human; cell adhesion and extracellular matrix protein; CADEC23;
KW neuroprotective; cytoskeletal; anorectic; immune disorder;
KW neurological disorder; developmental disorder;
KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
KW Tangier disease; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 36..260
FT /*tag= a
FT /product= "CADEC23"
XX
XX WO2004015396-A2.
XX
XX 19-FEB-2004.
PD
XX
XX 12-AUG-2003; 2003WO-US025418.
PF
XX
XX 13-AUG-2002; 2002US-0403781P.
PR

PR 30-AUG-2002; 2002US-0407034P.
PR 13-SEP-2002; 2002US-0410566P.
PR 24-SEP-2002; 2002US-0413482P.
PR 25-SEP-2002; 2002US-0413890P.
PR 08-NOV-2002; 2002US-0424904P.
PR 13-NOV-2002; 2002US-0426222P.
XX
XX (INCYTE) INCYTE CORP.
XX
XX Ellicott VS, Khare R, Emerling BM, Kahle AE, Tran UK, Jin P,
PI Becha SD, Margolis JP, Swarnakar A, Chawla NK, Ramkumar J,
PI Hatalla AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ,
PI Wang JT, Chien D, Yang YG;
XX
XX WPI; 2004-191795/18.
DR P-PSDB; ADM60794.
XX
XX New cell adhesion and extracellular matrix proteins, useful in
PT diagnosing, treating and preventing immune, neurological, developmental,
PT connective tissue and cell proliferative disorders including cancer.
XX
XX Claim 5; SEQ ID NO 65; 272bp; English.
XX
XX The present sequence encodes a human cell adhesion and extracellular
CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
CC cytoskeletal and anorectic activities. The CADECM polypeptides and
CC polynucleotides are useful in diagnosing, treating and preventing immune,
CC neurological, developmental, connective tissue and cell proliferative
CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and Tangle disease.
XX
XX Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;
SQ
Query Match 26.0%; Score 37.2; DB 12; Length 318;
Best Local Similarity 92.9%; Pred. No. 0.013;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 42
DB 47 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 88
RESULT 12
AAZ07020
ID AAZ07020 standard; DNA; 750 BP.
XX
XX AAZ07020;
AC
XX
XX 15-NOV-1999 (first entry)
DT
XX
XX Amelogenin X nucleotide sequence.
DE
XX
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KW multiple allelic site; apolipoprotein E; apoB; coronary artery disease;
KW Alzheimer's disease; ds.
XX
XX Unidentified.
OS
XX
XX WO940226-A2.
PN
XX
XX 12-AUG-1999.
PD
XX
XX 08-JAN-1999; 99WO-US000499.
PF
XX
XX 04-FEB-1998; 98US-00018595.
PR
XX
XX (PEKE) PERKIN-ELMER CORP.
PA
XX
XX Liyak KJ, Goodsaiaid F;
PI
XX
XX WPI; 1999-539985/45.
DR
XX
XX 5' nuclease amplification assay using fluorescence-quencher probes for

PT determination of a genotype at multiple allelic sites.
XX
XX Disclosure; Fig 10; 95pp; English.
PS
XX
XX The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apob genotypes. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for amelogenin X, which is
XX used in the exemplification of the present invention
SQ
Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;
Query Match 26.0%; Score 37.2; DB 2; Length 750;
Best Local Similarity 92.9%; Pred. No. 0.018;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATTTGTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTCTCTGGAGACGCTTTGCTATGCC 121
RESULT 13
AAZ07018
ID AAZ07018 standard; DNA; 793 BP.
XX
XX AAZ07018;
AC
XX
XX 15-NOV-1999 (first entry)
DT
XX
XX Amelogenin X nucleotide sequence.
DE
XX
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KW multiple allelic site; apolipoprotein E; apoB; coronary artery disease;
KW Alzheimer's disease; ss.
XX
XX Unidentified.
OS
XX
XX WO940226-A2.
PN
XX
XX 12-AUG-1999.
PD
XX
XX 08-JAN-1999; 99WO-US000499.
PF
XX
XX 04-FEB-1998; 98US-00018595.
PR
XX
XX (PEKE) PERKIN-ELMER CORP.
PA
XX
XX Liyak KJ, Goodsaiaid F;
PI
XX
XX WPI; 1999-539985/45.
DR
XX
XX 5' nuclease amplification assay using fluorescence-quencher probes for
PT determination of a genotype at multiple allelic sites.
PS
XX
XX Disclosure; Fig 8A; 95pp; English.
XX
XX The present invention describes first and second sets of fluorescer-

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 38.6217 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143
Sequence: 1 gattctgttgcctgcctgcctc.....gtttaagacagtaacagatc 143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfileseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.2	28.1	476	4 US-09-744-128-3	Sequence 3, Appl1
2	37.2	26.0	500	2 US-09-018-595B-3	Sequence 3, Appl1
3	37.2	26.0	500	3 US-09-324-709A-3	Sequence 3, Appl1
4	37.2	26.0	793	2 US-09-018-595B-1	Sequence 1, Appl1
5	37.2	26.0	793	3 US-09-324-709A-1	Sequence 1, Appl1
6	35.6	24.9	802	2 US-09-018-595B-2	Sequence 2, Appl1
7	35.6	24.9	802	3 US-09-324-709A-2	Sequence 2, Appl1
8	30	21.0	1284	4 US-09-328-352-2004	Sequence 2004, Ap
9	30	21.0	1230025	4 US-09-198-452A-1	Sequence 1, Appl1
10	29.6	20.7	116592	4 US-09-818-512-3	Sequence 1, Appl1
11	29.4	20.6	1431	3 US-09-054-782-1	Sequence 1, Appl1
12	29.4	20.6	1431	4 US-09-627-676-1	Sequence 1, Appl1
13	29.4	20.6	1431	4 US-09-054-272-15	Sequence 15, Appl1
14	29.4	20.6	1431	6 5466783-1	Patent No. 5466783
15	29.4	20.6	45175	3 US-09-453-702B-116	Sequence 116, App
16	29	20.3	3568	3 US-09-103-510B-1	Sequence 1, Appl1
17	29	20.3	3568	4 US-09-492-635A-1	Sequence 1, Appl1
18	29	20.3	3568	4 US-09-491-971A-1	Sequence 1, Appl1
19	29	20.3	3568	4 US-09-491-970A-1	Sequence 1, Appl1
20	28.6	20.0	886	4 US-09-270-767-31527	Sequence 31527, A
21	28.6	20.0	895	4 US-09-270-767-26797	Sequence 26797, A
22	28.6	20.0	1514	4 US-09-270-767-15199	Sequence 15199, A
23	28.6	20.0	1629	4 US-09-270-767-11232	Sequence 11232, A
24	28.2	19.7	1410	4 US-09-861-451A-13	Sequence 13, Appl1
25	27.8	19.4	792	4 US-09-710-279-203	Sequence 203, App
26	27.8	19.4	3841	4 US-09-710-279-3380	Sequence 3380, Ap
27	27.6	19.3	3666	4 US-09-710-279-4230	Sequence 4230, Ap

28	27.6	19.3	4066	4 US-09-710-279-4327	Sequence 4327, Ap
29	27.4	19.2	510	4 US-09-134-000C-106	Sequence 106, App
30	27	18.9	1138	4 US-09-270-767-26204	Sequence 26204, A
31	27	18.9	1669	4 US-09-591-466C-1	Sequence 1, Appl1
32	27	18.9	2123	4 US-09-270-767-10744	Sequence 10744, A
33	26.8	18.7	1914	4 US-09-248-796A-4911	Sequence 4911, Ap
34	26.8	18.7	5966	4 US-08-956-171E-22	Sequence 22, Appl1
35	26.8	18.7	5966	4 US-08-781-986A-22	Sequence 22, Appl1
36	26.6	18.6	9262	4 US-09-566-921-74	Sequence 74, Appl1
37	26.4	18.5	528	4 US-09-248-796A-6060	Sequence 6060, Ap
38	26.4	18.5	2070	4 US-09-540-236-1766	Sequence 1766, Ap
39	26.4	18.5	89047	4 US-09-596-002-34	Sequence 34, Appl1
40	26.2	18.3	804	3 US-09-134-001C-2205	Sequence 2205, Ap
41	26	18.2	936	4 US-09-489-039A-6691	Sequence 6691, Ap
42	26	18.2	1032	4 US-09-248-796A-8331	Sequence 8331, Ap
43	26	18.2	1511	3 US-08-847-065-18	Sequence 18, Appl1
44	26	18.2	1511	4 US-09-829-382-18	Sequence 18, Appl1
45	25.8	18.0	444	4 US-09-543-681A-1204	Sequence 1204, Ap

ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 2766/36983
; CURRENT APPLICATION NUMBER: US/09744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3
Query Match      28.1%; Score 40.2; DB 4; Length 476;
Best Local Similarity 93.3%; Pred. No. 6.8e-05;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY      1 GATTTGTTGCTGCTGCTGCTGAGACAGCTTGTATGCGGT 45
      59 GATCTGTTGCTGCTGCTGCTGAGACAGCTTGTATGCGGCT 103
RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSER: David J. Weitz,
; ADDRESSER: Wilson Somaini, Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-018-595B-3

Query Match 26.0%; Score 37.2; DB 2; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.0086;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTGGAGAGACGCTTGTGATGCC 42
Db 80 GATTTATTGCTGCTCCTGGAGAGACGCTTGTGATGCC 121

RESULT 3
US-09-324-709A-3
Sequence 3, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
TITLE OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-324-709A-3

Query Match 26.0%; Score 37.2; DB 3; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.0086;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTGGAGAGACGCTTGTGATGCC 42
Db 80 GATTTATTGCTGCTCCTGGAGAGACGCTTGTGATGCC 121

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1

Query Match 26.0%; Score 37.2; DB 2; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.0011;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GATTTGTTGCTGCTCCTGGAGAGACGCTTGTGATGCC 42
Db 80 GATTTATTGCTGCTCCTGGAGAGACGCTTGTGATGCC 121

RESULT 5
US-09-324-709A-1
Sequence 1, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David J. Weltz,
ADDRESSEE: Wilson Sonstini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-1
Query Match 26.0%; Score 37.2; DB 3; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.0011; 3; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GATTTGTTTGCCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTATTGCTGCTGCTGCTGAGAGAGCTTTGCTATGCC 121
RESULT 6
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 596223
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David J. Weltz,
ADDRESSEE: Wilson Sonstini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2
Query Match 24.9%; Score 35.6; DB 2; Length 802;
Best Local Similarity 90.5%; Pred. No. 0.0041; 4; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 GATTTGTTTGCCTGCTGCTGAGAGAGAGCTTTGCTATGCC 42
Db 80 GATTATTGCTGCTGCTGCTGAGAGAGCTTTGCTATGCC 121
RESULT 7
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David J. Weltz,
ADDRESSEE: Wilson Sonstini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362

```
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match
Best Local Similarity 90.5%; Pred. No. 0.0041;
Matches 38; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATTTGTTTGGCTGCTGCTGCGAGCAGCCTTGTATGCC 42
Db 80 GATTTGTTTGGCTGCTGCTGCGAGCAGCCTTGTATGCC 121

RESULT 8
US-09-328-352-2004/c
Sequence 2004, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 2004
LENGTH: 1284
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-2004

Query Match
Best Local Similarity 57.4%; Pred. No. 0.56;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 50 AAAATGCCCTGATATATTCATTTACAGCTTGAATTAAGTCCGCCACAG 109
Db 715 AGAAGAGCTTCACTACCTACCTTACCAAACTTGAAGCACTTTGCGCTTCAA 656

QY 110 TTGTAACCTTAGGGTTTAAGACAGTACAGATC 143
Db 655 TTGTAAGTTGATTTTTCAGCAACTTCAGACC 622

RESULT 9
US-09-198-452A-1/c
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)-(15000)
OTHER INFORMATION: n=a or c or g or t
```

```
NAME/KEY: misc_feature
LOCATION: (15001)-(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (30001)-(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (45001)-(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (60001)-(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (75001)-(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (90001)-(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (105001)-(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (120001)-(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (135001)-(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (150001)-(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (165001)-(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (180001)-(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (195001)-(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (210001)-(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (225001)-(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (240001)-(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (255001)-(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (270001)-(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (285001)-(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (300001)-(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (315001)-(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (330001)-(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (345001)-(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (360001)-(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

```
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
```

```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

Query Match 21.0%; Score 30; DB 4; Length 1230025;
Best Local Similarity 55.9%; Pred. No. 15;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 35 GCTATGCCCCGTGAGTAATAATACCCCTGATATATATTCATTTTCACAGAGTTGGAAATTA 94
Db 980791 GCTATTTCCTTTTTCATTTTCGATGCAATATTTCTTAAGTACTAAGTTTA 980732

Qy 95 AGTGTGCCCCACGAGTTGTAAGTTTAAAGACGTA 136
Db 980731 ATCCTTCTGCAATCTTCGAAACCATTTGGTTTAATAAGCA 980690

RESULT 10
US-09-818-512-3
Sequence 3, Application US/09818512
Patent No. 6537780
GENERAL INFORMATION:
APPLICANT: BEASLEY, Ellen et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
FILE REFERENCE: CL001192
CURRENT APPLICATION NUMBER: US/09/818,512
CURRENT FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 116592
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(116592)
OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

Query Match 20.7%; Score 29.6; DB 4; Length 116592;
Best Local Similarity 59.5%; Pred. No. 6.9; Mismatches 34; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 49 TAAATACCCCTGCATATATTCATTTTACACAGCTTGAAATTAAGTCCGCCACCA 108
Db 15953 TACATATCCCTTGATGATTTATTAATGTTCTTCATTAAGATTTGACTTTA 16012

Qy 109 GTTGTAACCTTAGGGTTAGAC 132
Db 16013 TTGATAGTTTATTAATTCAGTAC 16036

RESULT 11
US-09-054-782-1/c
; Sequence 1, Application US/09054782
; Patent No. 6171587
; GENERAL INFORMATION:
; APPLICANT: Mun, T.C.
; APPLICANT: Kretzmer, K.K.
; APPLICANT: Broze, Jr., G.J.
; TITLE OF INVENTION: DNA clone of human tissue
; TITLE OF INVENTION: factor inhibitor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Monsanto/Seale
; STREET: 700 Chesterfield Parkway No. 6171587th
; CITY: Chesterfield
; STATE: MO
; COUNTRY: USA
; ZIP: 63198
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/054,782
; FILING DATE: 03-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/463,602
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: 08/355,351
; FILING DATE: 13-DEC-1994
; APPLICATION NUMBER: 08/093,285
; FILING DATE: 15-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Dennis A
; REGISTRATION NUMBER: 34,547
; REFERENCE/DOCKET NUMBER: M 4941-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-737-6986
; TELEFAX: 737-6972
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; TISSUE TYPE: Human placenta
; IMMEDIATE SOURCE:
; LIBRARY: Human placental-derived lambda-P9 clone
; CLONE: Tissue factor inhibitor (TFI) cDNA
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1...132
; OTHER INFORMATION: *tag=a
; NAME/KEY: 3'UTR
; LOCATION: 1045...1431

OTHER INFORMATION: *tag=b
; NAME/KEY: CDS
; LOCATION: 3...104
; OTHER INFORMATION: *tag=c
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: EP 318451
; FILING DATE: 22-JUL-1988
; PUBLICATION DATE: 31-MAY-1988
; RELEVANT RESIDUES IN SEQ ID NO:
US-09-054-782-1

Query Match 20.6%; Score 29.4; DB 3; Length 1431;
Best Local Similarity 63.4%; Pred. No. 0.97; Mismatches 26; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 40 GCCCGTAGTAATACCCCTGCATATATTCATTTTACACAGCTTGAAATTAAGTCT 99
Db 1257 GCCAGTTAATTAATTAAGACCTAGAAATTAAGCAATTTAACAAGATTGAAAAGCACTCT 1198

Qy 100 GCCCCACCACT 110
Db 1197 GATACAAAGT 1187

RESULT 12
US-09-627-676-1/c
; Sequence 1, Application US/09627676
; Patent No. 6534276
; GENERAL INFORMATION:
; APPLICANT: Broze, George J.
; APPLICANT: Kretzmer, Kuniko K.
; APPLICANT: Mun, Tze-Chein
; TITLE OF INVENTION: Antibodies to human tissue factor
; TITLE OF INVENTION: inhibitor
; FILE REFERENCE: SM0941-10-US
; CURRENT APPLICATION NUMBER: US/09/627,676
; FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/054,782
; FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 08/463,323
; FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/355,351
; FILING DATE: 1993-07-15
; PRIOR APPLICATION NUMBER: 07/566,280
; FILING DATE: 1990-08-13
; PRIOR APPLICATION NUMBER: 07/123,753
; FILING DATE: 1987-11-23
; PRIOR APPLICATION NUMBER: 07/077,366
; FILING DATE: 1987-07-23
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: human
US-09-627-676-1

Query Match 20.6%; Score 29.4; DB 4; Length 1431;
Best Local Similarity 63.4%; Pred. No. 0.97; Mismatches 26; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 40 GCCCGTAGTAATACCCCTGCATATATTCATTTTACACAGCTTGAAATTAAGTCT 99
Db 1257 GCCAGTTAATTAATTAAGACCTAGAAATTAAGCAATTTAACAAGATTGAAAAGCACTCT 1198

Qy 100 GCCCCACCACT 110
Db 1197 GATACAAAGT 1187

RESULT 13
US-09-054-272-15/c
; Sequence 15, Application US/09054272

Patent No. 6692909
GENERAL INFORMATION:
APPLICANT: Lander, Eric S.
APPLICANT: Daley, George Q.
APPLICANT: Cargill, Michele
APPLICANT: Ireland, James S.
APPLICANT: Rozen, Steven G.
TITLE OF INVENTION: CODING SEQUENCE POLYMORPHISMS
TITLE OF INVENTION: IN VASCULAR PATHOLOGY GENES
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
City: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,272
FILING DATE: 01-Apr-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH198-05
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1431 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 133...1044
OTHER INFORMATION:
US-09-054-272-15

Query Match 20.6%; Score 29.4; DB 4; Length 1431;
Best Local Similarity 63.4%; Pred. No. 0.97;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 40 GCCCGTAGTAATAATCCCGCATATATTTCAATTTCAACAGCTTGAAATAAAGTCT 99
DB 1257 GCCAGTTAATAATTACAGACTAGATAAGCAATTTAACAAGATTGAAAGCAACTCT 1198
QY 100 GCCCCACCACT 110
DB 1197 GATACAAAGCT 1187

RESULT 14
546783-1/c
Patent No. 546783
APPLICANT: Mun, Tze-Chen, Kretzmer, Kuniko K., Broze, George J. Jr.
TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
NUMBER OF SEQUENCES: 26
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/93,285
FILING DATE: 15-JUL-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 566,280
FILING DATE: 13-AUG-1990
APPLICATION NUMBER: 123,753
FILING DATE: 23-NOV-1987
APPLICATION NUMBER: 77,366
FILING DATE: 23-JUL-1987
SEQ ID NO: 1:
LENGTH: 1431
546783-1

Query Match 20.6%; Score 29.4; DB 6; Length 1431;
Best Local Similarity 63.4%; Pred. No. 0.97;
Matches 45; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 40 GCCCGTAGTAATAATCCCGCATATATTTCAATTTCAACAGCTTGAAATAAAGTCT 99
DB 1257 GCCAGTTAATAATTACAGACTAGATAAGCAATTTAACAAGATTGAAAGCAACTCT 1198
QY 100 GCCCCACCACT 110
DB 1197 GATACAAAGCT 1187

RESULT 15
US-09-453-702B-116
Sequence 116, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Charles & Brady
STREET: 1 South Pinnacle Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 116:
SEQUENCE CHARACTERISTICS:
LENGTH: 45175
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 116:
US-09-453-702B-116

Query Match 20.6%; Score 29.4; DB 3; Length 45175;

Best Local Similarity 60.8%; Pred. No. 5.2;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 44 GTGAGTAAATATACCCCTGATATATTCATTTCAACAGCTTGAAATAAAAGTGTGCC 103

Db 82 GTTAGCATCAACCACTCTCAAAATATTCAGCACAACAGGTTAAATATACTGCACCC 141

QY 104 CACCAGTGGTAACTTAG 122

Db 142 GAACAATTGATTAACGACAG 160

Search completed: February 8, 2005, 22:26:15
Job time : 41.7328 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using BW model

Run on: February 8, 2005, 16:37:21 ; Search time 192.692 Seconds
(Without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 1 gattctgttcgtcgtctcc.....gtttaagacgtacaaatc 143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCF_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCFUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	75.8	53.0	600	US-10-027-632-287050	Sequence 287050,
C 2	75.8	53.0	600	US-10-027-632-287050	Sequence 287050,
C 3	74.6	52.2	600	US-10-027-632-287049	Sequence 287049,
C 4	74.6	52.2	600	US-10-027-632-287049	Sequence 287049,
5	74.6	52.2	259202	US-10-723-860-1311	Sequence 1311, Ap
6	40.2	28.1	65	US-09-908-975-1137	Sequence 1137, Ap
7	37.2	26.0	500	US-10-104-774-3	Sequence 3, Appli
8	37.2	26.0	500	US-10-455-150-3	Sequence 3, Appli
9	37.2	26.0	793	US-10-104-774-1	Sequence 1, Appli
10	37.2	26.0	793	US-10-455-150-1	Sequence 1, Appli
11	37.2	26.0	793	US-10-755-889-605	Sequence 605, App
12	35.6	24.9	802	US-10-104-774-2	Sequence 2, Appli

13	35.6	24.9	802	US-10-455-150-2	Sequence 2, Appli
14	35.6	24.9	852	US-10-723-860-5797	Sequence 5797, Ap
C 15	32.6	22.8	680	US-10-027-632-218969	Sequence 218969,
C 16	32.6	22.8	680	US-10-027-632-218969	Sequence 218969,
17	30.6	21.4	253861	US-10-741-601-5611	Sequence 5611, Ap
18	30.6	21.4	261817	US-10-087-193-2002	Sequence 2002, Ap
C 19	30.4	21.3	625	US-10-027-632-180446	Sequence 180446,
C 20	30.4	21.3	625	US-10-027-632-180447	Sequence 180447,
C 21	30.4	21.3	625	US-10-027-632-180448	Sequence 180448,
C 22	30.4	21.3	625	US-10-027-632-180446	Sequence 180446,
C 23	30.4	21.3	625	US-10-027-632-180447	Sequence 180447,
C 24	30.4	21.3	625	US-10-027-632-180448	Sequence 180448,
C 25	30.4	21.3	625	US-10-027-632-180446	Sequence 180446,
26	30	21.0	47804	US-10-292-796-629	Sequence 717, App
27	30	21.0	47804	US-10-292-796-629	Sequence 629, App
C 28	30	21.0	68140	US-10-087-192-1681	Sequence 1681, A
C 29	29.8	20.8	237	US-09-864-761-18974	Sequence 18974, A
C 30	29.8	20.8	520	US-10-264-049-2155	Sequence 2155, Ap
C 31	29.8	20.8	1513	US-09-938-842A-4427	Sequence 4427, Ap
C 32	29.8	20.8	1513	US-09-938-842A-4427	Sequence 4427, Ap
C 33	29.8	20.8	60327	US-10-052-482-187	Sequence 187, App
C 34	29.6	20.7	251	US-09-922-293-2795	Sequence 2795, Ap
C 35	29.6	20.7	116592	US-09-818-512-3	Sequence 3, Appli
36	29.6	20.7	116592	US-10-354-065-3	Sequence 3, Appli
37	29.4	20.6	741	US-10-198-846-9976	Sequence 9976, Ap
C 38	29.4	20.6	1431	US-10-377-817-1	Sequence 1, Appli
C 39	29.4	20.6	1431	US-10-886-556-1	Sequence 1, Appli
C 40	29.4	20.6	1612	US-10-198-846-9844	Sequence 9844, Ap
C 41	29.4	20.6	2849	US-09-880-107-2363	Sequence 2363, Ap
42	29.4	20.6	45175	US-10-114-170-116	Sequence 116, App
43	29.4	20.6	167343	US-09-962-436-281	Sequence 281, App
44	29.4	20.6	167343	US-09-962-436-281	Sequence 273, App
45	29.4	20.6	2140405	US-10-027-632-76212	Sequence 76212, A

ALIGNMENTS

RESULT 1
US-10-027-632-287050/c
; Sequence 287050, US20020198371A1
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027, 632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193, 483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185, 218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167, 363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156, 358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146, 002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050

Query Match 53.0%; Score 75.8; DB 13; Length 600;
Best local Similarity 78.5%; Pred. No. 2.1e-15;

```

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287049
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287049

Query Match
Best Local Similarity 78.5%; Score 74.6; DB 13; Length 600;
Matches 117; Conservative 0; Mismatches 24; Indels 8; Gaps 2

QY 1 GATTTGTTTGCCTGCTCTCTCGGAGACAGCCTTGTCTATGCCCGTAGTAAATACCCCT 60
    |||||
Db 245 GATTTATTTGGCTCCTCCTCGGAGAGAGCTTTTGATCATCTGTGAGTAAACACCCCT 186
    |||||

QY 61 -----GCATATATTTCATTTCAAGCTTGGAAATAAAGCTGCCACAGTTGCT 114
    |||||
Db 185 TGCATAGACGATGTCGCAATTTCCAAACTTGGACATATAAAATCTGCTC--ATAGTTGCT 128
    |||||

QY 115 AACTTAGGGTTTAAAGACGATCAAGATC 143
    |||||
Db 127 GAAATTAGGGTTTAAACAGTATGAGATC 99
    |||||

RESULT 4
US-10-027-632-287049/c
Sequence 287049, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287049
LENGTH: 600
TYPE: DNA
ORGANISM: Human

```



```

; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3
;
Query Match      26.0%; Score 37.2; DB 16; Length 500;
Best Local Similarity 92.9%; Pred. No. 0.032;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTGCTGGAGAGACCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTGCTGCTGGAGAGACCTTTTGCATGCC 121

RESULT 9
US-10-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1
;
Query Match      26.0%; Score 37.2; DB 13; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTGCTGGAGAGACCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTGCTGCTGGAGAGACCTTTTGCATGCC 121

RESULT 10
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
```

```

; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; Wilson Soncini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1
;
Query Match      26.0%; Score 37.2; DB 16; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038;
Matches 39; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTGCTGGAGAGACCTTTGCTATGCC 42
DB 80 GATTTATTTGCTGCTGCTGCTGGAGAGACCTTTTGCATGCC 121

RESULT 11
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 605
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-605
;
Query Match      26.0%; Score 37.2; DB 17; Length 793;
Best Local Similarity 92.9%; Pred. No. 0.038;
```


; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 218969
 ; LENGTH: 680
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-027-632-218969

Query Match 22.8%; Score 32.6; DB 13; Length 680;
 Best Local Similarity 60.9%; Pred. No. 1.4;
 Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy	14	TGCGTCTGGGAGAGCCTTGCTATGCGCGTGAATTAATACCCCTGCATATATTTCAA	73
Db	676	TGCGGCTGGGAGAGCCTTGCTATGCGCGTGAATTAATACCCCTGCATATATTTCAA	617
Qy	74	TTCACAAAGCTTGGAAATTAAGTCTG	100
Db	616	TGTGACAAAGCATGATATAATAAATATG	590

Search completed: February 9, 2005, 06:51:08
 Job time : 196.692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 1703.26 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-10

Perfect score: 143
1 gattctgttcgtcgtcctcc.....gtttaagacagtaagaatc 143

Sequence: IDENTITY_NUC

Scoring table: Gapop 10.0 ; Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41.8	29.2	303	7 W33932	W33932 mb54d02.r1
2	41.8	29.2	430	7 W40649	W40649 mc43h12.r1
3	41.8	29.2	536	7 W36345	W36345 mb72c12.r1
4	41.8	29.2	585	9 AY419441	AY419441 Mus muscu
5	41.8	29.2	645	6 CD773419	CD773419 AGENCOURT
6	41.8	29.2	707	6 CB056709	CB056709 NISC_j119
7	41.8	29.2	812	3 AK023358	AK023358 Mus muscu
8	41.8	29.2	843	6 CB588212	CB588212 AGENCOURT
9	41.8	29.2	859	6 CB588525	CB588525 AGENCOURT
10	41.8	29.2	865	6 CB589177	CB589177 AGENCOURT
11	41.8	29.2	886	6 CB587051	CB587051 AGENCOURT
12	41.8	29.2	889	6 CB589251	CB589251 AGENCOURT
13	41.8	29.2	891	6 CB574837	CB574837 AGENCOURT
14	41.8	29.2	913	6 CB590451	CB590451 AGENCOURT
15	41.8	29.2	919	6 CB587332	CB587332 AGENCOURT
16	41.8	29.2	928	6 CB590111	CB590111 AGENCOURT
17	41.8	29.2	945	2 BB614068	BB614068 BB614068
18	41.8	29.2	990	7 W12906	W12906 ma8p03.r1
19	41.8	29.2	1006	7 W29475	W29475 mb9p11.r1
20	41.8	29.2	1020	7 W08102	W08102 mb40g08.r1
21	40.2	28.1	217	7 R46913	R46913 Y140 Rat in
22	40.2	28.1	223	7 BB571643	BB571643 BB571643
23	40.2	28.1	449	7 R47024	R47024 Y350 Rat in
24	40.2	28.1	1031	6 CB575508	CB575508 AGENCOURT

25	39.2	27.4	373	7 R47135	R47135 Y722 Rat in
26	39.2	27.4	395	7 R47030	R47030 Y359 Rat in
27	39.2	27.4	556	7 R46934	R46934 Y169 Rat in
28	38.6	27.0	257	7 R47100	R47100 Y601 Rat in
29	38.2	26.7	303	7 R46903	R46903 Y124 Rat in
30	38.2	26.7	318	7 R47078	R47078 Y534 Rat in
31	38.2	26.7	457	7 R46947	R46947 Y186 Rat in
32	37.6	26.3	223	7 R47143	R47143 Y79 Rat inc
33	37.6	26.3	720	2 BB368600	BB368600 BB368600
34	37.6	26.3	1444	3 AK048251	AK048251 Mus muscu
35	37.2	26.0	429	6 CB473314	CB473314 sm68 E03
36	37.2	26.0	570	9 AY419439	AY419439 Homo sapi
37	37.2	26.0	570	9 AY419440	AY419440 Pan trogl
38	37	25.9	395	7 R46935	R46935 Y195 Rat in
39	36.6	25.6	478	7 R46933	R46933 Y166 Rat in
40	36.2	25.3	273	7 R47002	R47002 Y3 Rat inc1
41	36	25.2	1005	9 CL040441	CL040441 CH216-49N
42	35.6	24.9	286	7 R46887	R46887 Y106 Rat in
43	35.6	24.9	783	8 BG198114	BG198114 RST17499
44	35.2	24.6	513	8 A2725508	A2725508 RPT1-24-1
45	35	24.5	355	7 R46940	R46940 Y174 Rat in

ALIGNMENTS

RESULT 1
W33932
LOCUS
DEFINITION
W33932.1 GI:1315837
W33932
mb54d02.r1 Soares mouse p3NMf19.5 Mus musculus CDNA clone
IMAG:333219 5' similar to gb:M1005 Mouse amelogenin (MOUSE);,
mRNA sequence.

ACCESSION
W33932
VERSION
W33932.1 GI:1315837
KEYWORDS
EST.
SOURCE
Mus musculus
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gaisel,S., Kucaba,T., Lacey,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE
JOURNAL
COMMENT
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:214619
Seq primer: mob.REGA+ET
High quality sequence stop: 185.
Location/Qualifiers

FEATURES
source
1. 303
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:333219"
/dev stage="19.5 dpc total fetus"
/lab_host="PH108 (ampicillin resistant)"
/clone_lib="Soares mouse p3NMf19.5"
/note="Vector: pTR73D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCGAGTGGAGCGCGCATTTTATTTTATTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI

ORIGIN

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

Query Match 29.2%; Score 41.8; DB 7; Length 303;
Best Local Similarity 95.6%; Pred. No. 0.011;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 45
|||||
Db 81 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 125

RESULT 2
LOCUS W40649
DEFINITION m43h12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:351335 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.

ACCESSION W40649
VERSION W40649.1 GI:1324982
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 430)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schejlenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watsn.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:223335

Seq primer: EPrimer
High quality sequence stop: 323.
Location/Qualifiers

FEATURES

source

1..430
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:351335"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_1ib="Soares mouse p3NMF19.5"
/note="Vector: pRTT3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGAGCGGCGGAGCTTTGATGCGCGT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match

29.2%; Score 41.8; DB 7; Length 430;

Best Local Similarity 95.6%; Pred. No. 0.012;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 45
|||||
Db 69 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 113

RESULT 3
LOCUS W36345
DEFINITION mb72c12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);,
mRNA sequence.

ACCESSION W36345
VERSION W36345.1 GI:1318120
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 536)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schejlenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watsn.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:216366

Seq primer: EPrimer
High quality sequence stop: 359.
Location/Qualifiers

FEATURES

source

1..536
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:334966"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_1ib="Soares mouse p3NMF19.5"
/note="Vector: pRTT3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGAGCGGCGGAGCTTTGATGCGCGT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 29.2%; Score 41.8; DB 7; Length 536;
Best Local Similarity 95.6%; Pred. No. 0.012;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 45
|||||
Db 64 GATTTGTTGCTGCTGCTGAGAGAGAGCTTTGATGCGCGT 108

RESULT 4

AY419441 585 bp DNA linear GSS 17-DEC-2003
 LOCUS AY419441
 DEFINITION Mus musculus AMELX gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY419441
 VERSION AY419441.1 GI:39775398
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 585)
 Authors Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civejlo,D.R., Lu,F., Murphy,B.,
 Perriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
 Adams,M.D. and Cargill,M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 585)
 Authors Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civejlo,D.R., Lu,F., Murphy,B.,
 Perriera,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1..585
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>585
 /gene="AMELX"
 /locus_tag="HCM6891"
 ORIGIN
 Query Match 29.2%; Score 41.8; DB 9; Length 585;
 Best Local Similarity 95.6%; Pred. No. 0.013; 2; Indels 0; Gaps 0;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GATTTTGTTCCTGCTCCTCTGGAGCAGCCTTTGCTATGCCCT 45
 |||||||
 Db 12 GATTTTGTTCCTGCTCCTCTGGAGCAGCCTTTGCTATGCCCT 56
 |||||||
 RESULT 5 645 bp mRNA linear EST 02-JUL-2003
 CD773419
 LOCUS CD773419
 DEFINITION AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone
 IMAGE:30501267 5', mRNA sequence.
 ACCESSION CD773419
 VERSION CD773419.1 GI:32431921
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 645)
 NIH-MGC http://mgc.ncl.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Yoshihiko Yamada, Takashi Nakamura, NIDCR
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: NDCM23 row: k column: 04
 High quality sequence stop: 611.
 FEATURES
 source
 1..645
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30501267"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 190"
 /note="Organ: Pooled - Molar; Vector: pDNR-LIB; Site_1:
 Site 1 (ggccatgagcc); Site 2: Site 1 (ggccgctgagcc);
 Non-normalized full-length enriched library 5' and 3'
 adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATATGAGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.71
 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by
 PCR. This library was enriched for full-length clones and
 was constructed by Clontech Laboratories (Palo Alto, CA)
 Corp."

ORIGIN
 Query Match 29.2%; Score 41.8; DB 6; Length 645;
 Best Local Similarity 95.6%; Pred. No. 0.013;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GATTTTGTTCCTGCTCCTCTGGAGCAGCCTTTGCTATGCCCT 45
 |||||||
 Db 84 GATTTTGTTCCTGCTCCTCTGGAGCAGCCTTTGCTATGCCCT 128
 |||||||
 RESULT 6 707 bp mRNA linear EST 17-JAN-2003
 CB056709
 LOCUS CB056709
 DEFINITION NISC_j11905.w1 Soares NMBP13-15 Mus musculus cDNA clone
 IMAGE:4848584 5', mRNA sequence.
 ACCESSION CB056709
 VERSION CB056709.1 GI:27794996
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 707)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 cDNA Library Preparation: The I.M.A.G.E. Consortium/LNL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 info@image.llnl.gov
 Plate: LHAM10779 row: N column: 9
 Seq primer: T7 primer.
 FEATURES
 source
 1..707
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:4848584"
 /tissue_type="pituitary gland"
 /dev_stage="juvenile, 13-15 days"
 /lab_host="DH10B (phage-resistant)"

/clone.lib="Soares NMBP13-15"
 /note="Organ: brain; Vector: pT73D-Pac1; Site 1: NotI;
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-
 oligo (dT) primer
 5'-ACTGGAAGATTGCGCGCCGCTGTACCAAGTATTTT-3';
 double-stranded cDNA was ligated to EcoRI adaptors
 5'-ATTGGCAGCAGG-3' and 5'-CCTGCGCG-3' (Pharmacia),
 digested with NotI and cloned into the NotI and EcoRI
 sites of the pT73D-Pac1 vector. Library went through one
 round of normalization, and was constructed in the
 laboratory of M. Bento Soares (University of Iowa)."

ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 707;
 Best Local Similarity 95.6%; Pred. No. 0.013; 2; Indels 0; Gaps 0;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCGAGCAGCCTTGTATGCCCGT 45
 58 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 102

Db

RESULT 7 812 bp mRNA linear HTC 03-APR-2004
 AK029358 Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
 LOCUS library, clone:4833404E21 product:amelogenin, full insert sequence.
 ACCESSION AK029358
 VERSION AK029358.1 GI:26325321
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE Genome Res. 10 (10), 1617-1630 (2000)
 PUBMED 20499374
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 JOURNAL FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 REFERENCE 5
 AUTHORS Nature 409, 685-690 (2001)

TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
 JOURNAL Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 REFERENCE of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6
 (baee 1 to 812)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
 Hori, F., Imoto, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Konno, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku, Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE

Submitted (16-JUN-2001) Yoshinori Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.jp/
 URL: http://fantom.gsc.riken.jp/
 Location/Qualifiers

FEATURES

source

1..812
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM.DB:4833404E21"
 /db_xref="taxon:10090"
 /clone="4833404E21"
 /issue_type="head"
 /clone.lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 73..663
 /note="unnamed protein product; amelogenin (MGD) [MGI:88005,
 GB|D11768, evidence: BLASTN, 99%, match=799]
 putative"

CDS

/codon_start=1
 /protein_id="BAC26415.1"
 /db_xref="GI:26325322"
 /translation="MTWILFACLLGAAPMPHPHSPGYINISVLTPLPKYQS
 MIROPYPSYGYEPMGWLHQIIPVLSQAPHPSHLQPHNLPPVPAQAPVAPQPMQ
 PVPGHSMTPYOHQPNIPSPSAQPFQCPQPOAIIPQSHQPMQPSLHPMOPAPQ
 PPLPLPFSMOPLSPLPLPLPLPAMPATDKTKREVD"
 793..798
 /note="putative"
 812
 /note="putative"

polyA_site

812
 /note="putative"

ORIGIN

Query Match 29.2%; Score 41.8; DB 3; Length 812;
 Best Local Similarity 95.6%; Pred. No. 0.014; 2; Indels 0; Gaps 0;
 Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTGTTGCTGCTGCGAGCAGCCTTGTATGCCCGT 45
 84 GATTTGTTGCTGCTGCTGCGAGCAGCCTTGTATGCCCGT 128

Db

RESULT 8 843 bp mRNA linear EST 03-APR-2003
 CB588212
 LOCUS CB588212
 DEFINITION AGENCOURT 12771359 NIH MGC 136 Mus musculus cDNA clone
 IMAGE:30293239 5', mRNA sequence.
 ACCESSION CB588212
 VERSION CB588212.1 GI:29506068
 KEYWORDS EST.

SOURCE
ORGANISM Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM338 row: 9 column: 08
High quality sequence stop: 636.
Location/Qualifiers
1. 843
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293239"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCGCC(T)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 843;
Best Local Similarity 95.6%; Pred. No. 0.014; 2; Indels 0; Gaps 0;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTTGTTCCTGCTCCTCGGAGACGCTTTGCTATGCCCT 45
|||||
Db 51 GATTTTGTTCCTGCTCCTCGGAGACGCTTTGCTATGCCCT 95
|||||

RESULT 9
LOCUS CB588525 859 bp mRNA linear EST 03-APR-2003
DEFINITION AGENCOURT 12567944 NIH MGC 136 Mus musculus CDNA clone
IMAGE:30289751 5', mRNA sequence.
CB588525
CB588525.1 GI:29506381
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: NDAM329 row: e column: 24
High quality sequence stop: 763.
Location/Qualifiers
1. 859
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCCGCC(T)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 29.2%; Score 41.8; DB 6; Length 859;
Best Local Similarity 95.6%; Pred. No. 0.014; 2; Indels 0; Gaps 0;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GATTTTGTTCCTGCTCCTCGGAGACGCTTTGCTATGCCCT 45
|||||
Db 70 GATTTTGTTCCTGCTCCTCGGAGACGCTTTGCTATGCCCT 114
|||||

RESULT 10
LOCUS CB589177 865 bp mRNA linear EST 03-APR-2003
DEFINITION AGENCOURT 12770446 NIH MGC 136 Mus musculus CDNA clone
IMAGE:30290534 5', mRNA sequence.
CB589177
CB589177.1 GI:29507033
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM331 row: f column: 15
High quality sequence stop: 630.
Location/Qualifiers
1. 865
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290534"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw

TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDM342 row: d column: 24
High quality sequence start: 19
High quality sequence stop: 658.
Location/Qualifiers
1. .891
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30294719"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTTACATCGCGCGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 29.2%; Score 41.8; DB 6; Length 891;
Best Local Similarity 95.6%; Pred. No. 0.014;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GATTTGTTGCTGCTGCTGAGAGCGCTTGTATGCCCCCT 45
Db 90 GATTTGTTGCTGCTGCTGAGAGCGCTTGTATGCCCCCT 134

RESULT 14
CBS90451 913 bp mRNA linear EST 03-APR-2003
AGENCOURT_12770290 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30286944 5', mRNA sequence.
CBS90451
VERSION CBS90451.1 GI:29508307
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 913)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDM342 row: a column: 01
High quality sequence start: 12
High quality sequence stop: 640.
Location/Qualifiers

FEATURES
source

source
1. .913
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30286944"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTTACATCGCGCGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 29.2%; Score 41.8; DB 6; Length 913;
Best Local Similarity 95.6%; Pred. No. 0.014;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GATTTGTTGCTGCTGCTGAGAGCGCTTGTATGCCCCCT 45
Db 82 GATTTGTTGCTGCTGCTGAGAGCGCTTGTATGCCCCCT 126

RESULT 15
CBS87332 919 bp mRNA linear EST 03-APR-2003
AGENCOURT_12971606 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30291971 5', mRNA sequence.
CBS87332
VERSION CBS87332.1 GI:29505188
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 919)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: NDM335 row: b column: 12
High quality sequence start: 216
High quality sequence stop: 325.
Location/Qualifiers
1. .919
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30291971"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTTACATCGCGCGCCGCC(T)15-3', size selected for the >1kb fragments, average insert size

FEATURES
source

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 / Search time 814.048 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-11

Sequence: 1 gattctgtctgcctgcctcc.....gttaagacagctaacagatc 146

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb ba: *
2: gb hntg: *
3: gb ln: *
4: gb om: *
5: gb ov: *
6: gb pac: *
7: gb ph: *
8: gb pl: *
9: gb pr: *
10: gb ro: *
11: gb str: *
12: gb sy: *
13: gb un: *
14: gb vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.2	46.0	5712	4	AB091793 Equus cab
2	64	43.8	6451	4	AB091789 Bos tauru
3	56.6	38.8	6931	4	AB091792 Sus scrof
4	54.8	37.5	6465	9	AB091781 Pan trogl
5	54.8	37.5	7425	4	AB091791 Sus scrof
6	54.4	37.3	5591	4	AB091794 Equus cab
7	53.6	36.7	6442	9	AB091783 Salmi
8	53.6	36.7	8004	9	AB091786 Lemur cat
9	52.8	36.2	5562	9	AB091787 Ootlemur
10	52.2	35.8	8810	9	AY040206 Homo sapi
11	52.2	35.8	158142	2	AL357130 Homo sapi
12	52.2	35.8	259202	2	AC002366 Human Xp2
13	51.4	34.2	5684	9	AB091785 Lemur cat
14	50	34.2	695	11	BV089295 RPAMSE00
15	50	34.2	695	11	BV097603 RPAMSE00
16	49.8	34.1	463	10	D8306362 Mus muscu
17	49.8	34.1	9384	10	AF294397 Mus muscu
18	49.8	34.1	95826	10	AL805974 Mouse DNA
19	49.2	33.7	363	11	BV089294 RPAMSE00

C 20	49.2	33.7	363	11	BV097602	BV097602 RPAMSE00
C 21	49	33.6	6264	4	AB091790	AB091790 Bos tauru
C 22	47.6	32.6	212886	2	AC093946	AC093946 Rattus no
C 23	47.6	32.6	234471	2	AC121424	AC121424 Rattus no
C 24	42.8	29.3	7163	9	AB091782	AB091782 Pan trogl
C 25	42.8	29.3	38765	9	BS000568	BS000568 Pan trogl
C 26	42.8	29.3	177654	2	AP000918	AP000918 Homo sapi
C 27	42.8	29.3	177726	9	AC145770	AC145770 Pan trogl
C 28	42.8	29.3	150089	9	BS000576	BS000576 Pan trogl
C 29	42.8	29.3	200214	9	AC013412	AC013412 Homo sapi
C 30	40.2	27.5	746	4	AF215890	AF215890 Capra hir
C 31	40.2	27.5	752	4	AF215889	AF215889 Capra hir
C 32	40.2	27.5	757	4	BOVAMLGNTI	BOVAMLGNTI Bovine (cla
C 33	40.2	27.5	820	4	BOVAMLGNTI	BOVAMLGNTI Bovine (cla
C 34	38.6	26.4	778	4	SSU43405	SSU43405 Sus scrofa
C 35	38	26.0	159826	5	AL831730	AL831730 Zebrafish
C 36	38	26.0	321335	2	BX004986	BX004986 Dario rer
C 37	37.8	25.9	168866	2	AC136391	AC136391 Rattus no
C 38	37.8	25.9	216837	2	AC126597	AC126597 Rattus no
C 39	37.8	25.9	248167	2	AC106455	AC106455 Rattus no
C 40	37.4	25.6	106506	8	AP006137	AP006137 Lotus cor
C 41	37	25.3	7454	9	AB091784	AB091784 Salmi
C 42	37	25.3	92896	9	AL603908	AL603908 Human DNA
C 43	37	25.3	150941	2	BX957281	BX957281 Dario rer
C 44	36.8	25.2	104995	9	BS000008	BS000008 Pan trogl
C 45	36.8	25.2	123600	9	BS000007	BS000007 Pan trogl

ALIGNMENTS

RESULT 1
AB091793
LOCUS
DEFINITION
EQUUS caballus AMELX gene for amelogenin, partial cds.
ACCESSION
AB091793
VERSION
AB091793.1 GI:29126038
KEYWORDS
SOURCE
ORGANISM
EQUUS caballus (horse)
EQUUS caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

REFERENCE
1
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Inai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
JOURNAL
MEDLINE
22608569
PUBMED
12672962
REFERENCE
2 (bases 1 to 5712)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Soken), Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
Location/Qualifiers
1..5712
/organism="EQUUS caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/sex="male"
174..5712
/gene="AMELX"
join(174..233,1578..1589)
/gene="AMELX"
join(1590..1643,3566..3613,4993..5037,5305..55712)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC6112.1"
/db_xref="GI:29126039"

FEATURES

source

gene
5'UTR
CDS

Db 2119 GATTTCTTTCCTGCTGCTTTTGGAGAGCATCCCTGTATGCTGCTGATTAAG--CTA 2176

Qy 61 TTTCATTTTCAAGCTTGAATAAATCTGCTCAGA 101

Db 2177 AGTCAATTTCAATGCTTGGAATAAATTCACACATA 2217

RESULT 4
AB091781 6465 bp DNA linear PRI 02-MAY-2003
LOCUS Pan troglodytes AMELX gene for amelogenin, partial cds.
DEFINITION
AB091781
VERSION
AB091781.1 GI:29126015
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

REFERENCE
AUTHORS
TITLE
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
JOURNAL
MEDLINE
22608569
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

REFERENCE
AUTHORS
TITLE
2 (bases 1 to 6465)
Iwase, M., Saita, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kikusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@min@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES
source
Location/Qualifiers
1. 6465
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/sex="male"
927. 6465
/gene="AMELX"
join(927. 982, 2283. 2294)
/gene="AMELX"
join(2295. 2348, 4284. 4331, 5607. 5648, 5740. 5781,
6052. 6465)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66101.1"
/db_xref="GI:29126015"
/translation="MGTWILFACLLGAAPMP.PPHGPHGYINFSYENSHQAINVD
RTALVPLKMYQSIIRPYSYGEBPMGMLHQLIPVLSQOHPTTHLQPHHILPM
VPAQOPVLPQOPMPLPGQHSMTPTIOHNPMLPPPAQOPVQPOVQPOPHQPOPPV
HPMQPLPQPPPLPMPFMPQPLPMLPLTLLEAMPSTDKR"

ORIGIN
Query Match 37.5%; Score 54.8; DB 9; Length 6465;
Best Local Similarity 77.8%; Pred. No. 0.00015;
Matches 84; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

Qy 1 GATTTGTTTCCGCTGCTGCTGAGAGAGAGCTTCAGTATGCTGCTGATAAATTTTC--- 57
Db 2306 GATTTTATTTGCTGCTGCTGCTGAGAGAGAGCTTTTGCATGCTGCTGATAAATTCACACCC 2365

Qy 58 -----CAATTTCAATTTCAAGAGCTTGGAATAAATCTGCC 96
Db 2366 TGCATAGTCAGTGTCAATTTCAAGAGCTTGGAATAAATCTGCC 2413

RESULT 5
AB091791 7425 bp DNA linear MAM 02-MAY-2003
LOCUS Sus scrofa AMELX gene for amelogenin, partial cds.
DEFINITION

ACCESSION
VERSION
AB091791.1 GI:29126034
KEYWORDS
SOURCE
Sus scrofa (pig)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

REFERENCE
AUTHORS
TITLE
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
JOURNAL
MEDLINE
22608569
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

REFERENCE
AUTHORS
TITLE
2 (bases 1 to 7425)
Iwase, M., Saita, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kikusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@min@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES
source
Location/Qualifiers
1. 7425
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/sex="male"
1491. 7425
/gene="AMELX"
join(1491. 1548, 3140. 3151)
/gene="AMELX"
join(3152. 3205, 5316. 5363, 6619. 6660, 6717. 6761,
7033. 7425)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66110.1"
/db_xref="GI:29126035"
/translation="MGTWILFACLLGAAPMP.PPHGPHGYINFSYEDLYLRAIRD
RTAFVLPKMYQNMIRHPTSYGEBPMGMLHQLIPVLSQOHPTTHLQPHHILPM
VPAQOPVLPQOPMPLPGQHSMTPTIOHNPMLPPPAQOPVQPOVQPOPHQPOPPV
HPMQPLPQPPPLPMPFMSQSLPDLPLKMPAT"

ORIGIN
Query Match 37.5%; Score 54.8; DB 4; Length 7425;
Best Local Similarity 75.0%; Pred. No. 0.00015;
Matches 87; Conservative 0; Mismatches 17; Indels 12; Gaps 1;

Qy 1 GATTTGTTTCCGCTGCTGCTGAGAGAGAGCTTCAGTATGCTGCTGATAAATTTTCGA 60
Db 3163 GATTTGTTTCCGCTGCTGCTGAGAGAGAGCTTCAGTATGCTGCTGATAAATTCACACCC 3222

Qy 61 TT-----TCAATTTCAAGAGCTTGGAATAAATCTGCCGATTT 104
Db 3223 TTCTACGTCATGCTCAATGCTCATACGCTTGGAATAAATCTACCCCGAGTT 3278

RESULT 6
AB091794 5591 bp DNA linear MAM 02-MAY-2003
LOCUS Equus caballus AMELX gene for amelogenin, partial cds.
DEFINITION
AB091794
VERSION
AB091794.1 GI:29126040
KEYWORDS
SOURCE
ORGANISM
Equus caballus (horse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.

REFERENCE
AUTHORS
TITLE
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species

5' UTR join(1382..1437,3040..3051)
/gene="AMELY"
/join(3052..3105,5192..5239,6461..>6505)
/gene="AMELY"
/codon_start=1
/product="amelogenin"
/protein_id="BAC6106.1"
/db_xref="GI:29126026"
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS
MIRPP"

ORIGIN

Query Match 36.7%; Score 53.6; DB 9; Length 8004;
Best Local Similarity 78.6%; Pred. No. 0.00032;
Matches 77; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 GATTTGTTGCTGCTGCTGCTGAGAGAGCCTTCAGTATGCTGTGATAAATTTCCAA 60
|||||
DB 3063 GATTTGTTGCTGCTGCTGCTGAGAGAGCCTTTGTATGCTGTGATAAATTTCCCT 3122
|||||

QY 61 TTTCACAGCTTGGAAATTAATTTGCC 96
|||||
DB 3123 TGCCTAAGCTGTAAAGCTTAATTAATTAATTTGCC 3160
|||||

RESULT 9
AB091787 5562 bp DNA linear PRI 02-MAY-2003
LOCUS AB091787
DEFINITION Otollemur garnettii AMELX gene for amelogenin, partial cds.
ACCESSION AB091787
VERSION AB091787.1 GI:29126027
KEYWORDS Otollemur garnettii (small-eared galago)
SOURCE Otollemur garnettii
ORGANISM Otollemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otollemur.
REFERENCE 1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
AUTHORS From the cover: The amelogenin loci span an ancient pseudautosomal
TITLE boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569
PUBMED 12672962
REFERENCE 2 (bases 1 to 5562)
AUTHORS Iwase, M., Saita, Y. and Takahata, N.
TITLE Direct Submision
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
ADVANCED STUDIES (Soken), Department of Biosystems Science;
Shonan Kokuai-mura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@nkc.kyushu-u.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
FEATURES
source Location/Qualifiers
1..5562
/organism="Otollemur garnettii"
/mol_type="genomic DNA"
/db_xref="taxon:30611"
/sex="male"
267..5562
/gene="AMELX"
/join(267..321,1401..1412)
/gene="AMELX"
/join(1413..1466,3385..3432,4699..4740,4832..4876,
5125..>5562)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC6107.1"
/db_xref="GI:29126028"
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS
MIRPP"

ORIGIN

ORIGIN

Query Match 36.2%; Score 52.8; DB 9; Length 5562;
Best Local Similarity 69.6%; Pred. No. 0.00053;
Matches 110; Conservative 0; Mismatches 27; Indels 21; Gaps 2;

QY 1 GATTTGTTGCTGCTGCTGCTGAGAGAGCCTTCAGTATGCTGTGATAAATTTCC--- 57
|||||
DB 1424 GATTTGTTGCTGCTGCTGCTGAGAGAGCTTTTGCATGCCGTGATAAATTTCCCT 1483
|||||

QY 58 -----CAATTCATTTCAACAAGCTTGAATAAATTCGCTCAGATTTTAT 108
|||||
DB 1484 TGCATAAGCTAGTATCAATTTTCAATGCTTAAGATAAATAATCTGCTCAACAGTT---- 1539
|||||

QY 109 TTAAGGTTAATTTAAGGTTTAAGACATCAAGATC 146
|||||
DB 1540 -----GGTAACTTCAGGATTTAAACATGATGATC 1572
|||||

RESULT 10
AY040206 8810 bp DNA linear PRI 10-APR-2002
LOCUS AY040206
DEFINITION Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
ACCESSION AY040206
VERSION AY040206.1 GI:15028582
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 8810)
AUTHORS Hart, P.S., Hart, T.C., Simer, J.P. and Wright, J.T.
TITLE A nomenclature for X-linked amelogenesis imperfecta
JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)
MEDLINE 21920287
PUBMED 11922868
REFERENCE 2 (bases 1 to 8810)
AUTHORS Hart, S., Hart, T.C. and Wright, J.T.
TITLE Direct Submision
JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
3550 Terrace St., 572A Scalle Hall, Pittsburgh, PA 15090, USA
FEATURES
source Location/Qualifiers
1..8810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.3-p22.1"
1..8791
/gene="AMELX"
1..1416
/gene="AMELX"
1386..1391
/gene="AMELX"
/join(1417..1472,2781..2846,4783..4830,6240..6281,
6552..6977,8611..8791)
/gene="AMELX"
/product="amelogenin precursor"
1417..1472
/gene="AMELX"
/number=1
2781..2846
/gene="AMELX"
/number=2
/join(2793..2846,4783..4830,6240..6281,6552..6977,
8611..8616)
/gene="AMELX"
/codon_start=1
/product="amelogenin precursor"
/protein_id="AAK77213.1"
/db_xref="GI:15028583"
/translation="MGTWILFACLLGAFAFMPLOPHRGHGYINFSYEVLTPLKWTOS
MIRPP"

ORIGIN

MEPMQPLPMLPDLTLEAMPSTDKREVD"

sig_peptide 2793..2840
 /gene="AMELX"
 exon 4783..4830
 /gene="AMELX"
 /number=3
 exon 6107..6148
 /gene="AMELX"
 /number=4
 exon 6240..6281
 /gene="AMELX"
 /number=5
 exon 6552..6977
 /gene="AMELX"
 /number=6
 exon 8611..8791
 /gene="AMELX"
 /number=7
 exon 8785..8791
 /gene="AMELX"

polyA_signal

ORIGIN

Query Match 35.8%; Score 52.2; DB 9; Length 8810;
 Best Local Similarity 76.6%; Pred. No. 0.00076;
 Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;

Oy 1 GATTTGTTGCTGCGCTCCGAGGAGCGCTTCAGTATGCTGTGATAAATTC--- 57
 |||||
 Db 2804 GATTATATTCCTCCCTCCGAGGAGCGCTTTCAGTATGCTGTGATAAACAACCCCT 2863
 |||||

Oy 58 -----CAATTCGAATTCACAACTTGGAATTAATAATCTGC 95
 |||||
 Db 2864 TGCATAGTCAGTGTCCCAATTCACAACTTGGAATTAATAATCTGC 2910
 |||||

RESULT 11
 AL357130/c 158142 bp DNA linear HTG 10-JUN-2001
 LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.
 AL357130.3 GI:9863814
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 McIay, K.
 Direct Submission
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 21, 2000 this sequence version replaced gi:9714076.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: bbi69E15
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 155448 bases at least Q40
 Consensus quality: 156466 bases at least Q30
 Consensus quality: 157038 bases at least Q20
 Insert size: 157842; sum-of-contigs
 Insert size: 160705; 33.4% error; agarose-fp
 Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality
 coverage: 4.56x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1..158142
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone_1fb="RP13-169E15"
 /clone_1lb="RP13-13.1"
 1..68897
 /note="assembly fragment: 01108
 fragment chain:1
 clone_end:17
 vector_side:left"
 68998..122842
 /note="assembly fragment: 00101
 fragment chain:1"
 122943..129584
 /note="assembly fragment: 00361
 fragment chain:1"
 129685..158142
 /note="assembly fragment: 01285
 fragment chain:1
 clone_end:SP6
 vector_side:right"

misc_feature
 /note="assembly fragment: 01108
 fragment chain:1
 clone_end:17
 vector_side:left"
 68998..122842
 /note="assembly fragment: 00101
 fragment chain:1"
 122943..129584
 /note="assembly fragment: 00361
 fragment chain:1"
 129685..158142
 /note="assembly fragment: 01285
 fragment chain:1
 clone_end:SP6
 vector_side:right"

misc_feature
 /note="assembly fragment: 01285
 fragment chain:1
 clone_end:SP6
 vector_side:right"

misc_feature
 /note="assembly fragment: 01285
 fragment chain:1
 clone_end:SP6
 vector_side:right"

ORIGIN

Query Match 35.8%; Score 52.2; DB 2; Length 158142;
 Best Local Similarity 76.6%; Pred. No. 0.00069;
 Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;

Oy 1 GATTTGTTGCTGCGCTCCGAGGAGCGCTTCAGTATGCTGTGATAAATTC--- 57
 |||||
 Db 44314 GATTATATTCCTCCCTCCGAGGAGCGCTTTCAGTATGCTGTGATAAACAACCCCT 44255
 |||||

Oy 58 -----CAATTCGAATTCACAACTTGGAATTAATAATCTGC 95
 |||||
 Db 44254 TGCATAGTCAGTGTCCCAATTCACAACTTGGAATTAATAATCTGC 44208
 |||||

RESULT 12
 AC002366 259202 bp DNA linear PRI 11-JUN-1998
 LOCUS Human XP22 BAC CT-285115 (from CalTech/Research Genetics), PAC
 DEFINITION RPI11-27C22 (from Roswell Park Cancer Center), and Cosmid U5B5
 (from Lawrence Livermore), complete sequence.
 AC002366 U79549 U70036
 VERSION
 AC002366.1 GI:2739349
 HTG.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 259202)
 Muzny, D., Ansari-Lari, M.A., Timme, K.M., Yu, W., Dugan, S., Lu, J.,
 Shen, Y., Rowland, K., Liu, W., Perez, L., Ding, Y., Haywood, M.,
 Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,
 Worley, K., Chen, E., Forcum, J., Atkinson, A.D., Chiu, M.W.,
 Gorrell, J.H., Brundage, E., Di, W., Chnault, C., Nelson, D. and
 Gibbs, R.A.

REFERENCE	TITLE	JOURNAL
REFERENCE	Direct Submission	Unpublished
REFERENCE	2 (bases 1 to 259202)	
REFERENCE	Chiu,M.W.	
REFERENCE	Direct Submission	
REFERENCE	Submitted (23-JUL-1997)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 259202)	
REFERENCE	Chiu,M.W.	
REFERENCE	Direct Submission	
REFERENCE	Submitted (02-JAN-1998)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 259202)	
REFERENCE	Chiu,M.W.	
REFERENCE	Direct Submission	
REFERENCE	Submitted (11-JUN-1998)	Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	On Jan 2, 1998 this sequence version replaced gi:2642176	
REFERENCE	Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.	
REFERENCE	The repeat regions shown were identified using RepeatMasker by Adrian Smt.	
REFERENCE	Sequence similarities were identified using Powerblast by Jinghui Zhang.	
REFERENCE	Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.	
REFERENCE	location/Qualifiers	
REFERENCE	1. 259202	
REFERENCE	/organism="Homo sapiens"	
REFERENCE	/mol_type="genomic DNA"	
REFERENCE	/db_xref="taxon:9606"	
REFERENCE	/chromosome="X"	
REFERENCE	/map="Xp22"	
REFERENCE	/clone="CT-285115, RPTC11-27C22, U35B5"	
REFERENCE	/complement(1. 147)	
REFERENCE	/note="Overlap 1. 147 in AC004467"	
REFERENCE	35. 127	
REFERENCE	/rpt_family="MIR"	
REFERENCE	/complement(1087. 1502)	
REFERENCE	/rpt_family="1"	
REFERENCE	/complement(5292. 5579)	
REFERENCE	/rpt_family="Alu"	
REFERENCE	5302. 5553	
REFERENCE	/rpt_family="SVA"	
REFERENCE	/complement(6009. 6063)	
REFERENCE	/rpt_family="11MC2"	
REFERENCE	/complement(7092. 7382)	
REFERENCE	/rpt_family="L1MS"	
REFERENCE	7423. 7721	
REFERENCE	/rpt_family="Alu"	
REFERENCE	9771. 10072	
REFERENCE	/rpt_family="11MC2"	
REFERENCE	/complement(10072. 11146)	
REFERENCE	/rpt_family="THR"	
REFERENCE	/complement(11156. 11528)	
REFERENCE	/rpt_family="THE1b"	
REFERENCE	/complement(12944. 13151)	
REFERENCE	/rpt_family="MER20"	
REFERENCE	13862. 14010	
REFERENCE	/rpt_family="MLT1f"	
REFERENCE	14063. 14352	
REFERENCE	/rpt_family="MERA2a"	
REFERENCE	15847. 16170	
REFERENCE	/rpt_family="Alu"	
REFERENCE	/complement(18743. 19117)	
REFERENCE	/rpt_family="MER7"	

repeat_region	23507..23614	/rpt family="L1"
repeat_region	23615..24425	/rpt family="L1PA11"
repeat_region		complement(25475..25968)
STS		/rpt family="MLT1d"
	21234..27683	
	/note="similar to EST with GenBank Accession Number H48827"	
repeat_region	/db xref="dbSTS:G28372"	
repeat_region	complement(28186..28366)	
repeat_region	/rpt family="MER20"	
repeat_region	28715..28995	
repeat_region	/rpt family="Alu"	
repeat_region	complement(30217..30408)	
repeat_region	/rpt family="L1MA5"	
repeat_region	31195..31364	
repeat_region	/rpt family="Alu"	
repeat_region	complement(33939..34460)	
repeat_region	/rpt family="L1MA2"	
repeat_region	34952..35384	
repeat_region	/rpt family="MLT1d"	
repeat_region	36035..36126	
repeat_region	/rpt family="MIR2"	
repeat_region	36850..37035	
repeat_region	/rpt family="MER20"	
STS	38042..38250	
	/db xref="dbSTS:L47508"	
repeat_region	38122..38510	
repeat_region	/rpt family="MSTa"	
repeat_region	38797..40103	
repeat_region	/rpt family="MIR"	
repeat_region	complement(42465..42539)	
repeat_region	/rpt family="MIR"	
repeat_region	47139..47335	
repeat_region	/rpt family="MER3"	
repeat_region	complement(48993..49166)	
repeat_region	/rpt family="MIR"	
repeat_region	complement(49718..49918)	
repeat_region	/rpt family="L1PA11"	
repeat_region	complement(51943..52505)	
repeat_region	/rpt family="L1MB3"	
repeat_region	53226..53677	
repeat_region	/rpt family="MLT1d"	
repeat_region	complement(54846..54966)	
repeat_region	/rpt family="MIR"	
repeat_region	55040..55226	
repeat_region	/rpt family="MER20"	
repeat_region	55691..55837	
repeat_region	/rpt family="Alu"	
repeat_region	56051..56486	
repeat_region	/rpt family="L1ME3a"	
repeat_region	58557..59030	
repeat_region	/rpt family="MLT1d"	
repeat_region	complement(59270..59428)	
repeat_region	/rpt family="MER5"	
repeat_region	complement(59847..60030)	
repeat_region	/rpt family="MER20"	
repeat_region	complement(61290..61384)	
repeat_region	/rpt family="L1PBL"	
repeat_region	61683..61811	
repeat_region	/rpt family="MER20"	
repeat_region	61907..62113	
repeat_region	/rpt family="L1MB7"	
repeat_region	complement(63588..63969)	
repeat_region	/rpt family="L1MA10"	
repeat_region	64800..65073	
repeat_region	/rpt family="Alu"	
repeat_region	complement(67157..67209)	
repeat_region	/rpt family="MIR"	
gene	67637..75845	
	/gene="HAMEIX"	
chromoter	67637..68496	

Studies (Sokendai), Department of Biosystems Science;

/note="SNPs developed from assay sequences derived from 15

ORIGIN

different strains-of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, -129/Sv, AKR/J, B10.D2-H2/oshJ, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, -CAST/Ei, DBA/2J, NZB/MPJ, NZB/BlmJ, NZM/Lac, SPRET/Ei,..."

<1..>695

Query Match 34.2%; Score 50; DB 11; Length 695;

Best Local Similarity 68.2%; Pred. No. 0.0032; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Query 61 TT-----TCCAAATTTCACAGCTTGGAAATTAATAATCTGCTCAGATTTTATT 109
 419 TTACTAATTCATCCAAATTTGATATCTTGGAACTAGAAATCTGCTCCCAAGTT----- 364

Query 110 TAGGGTTAAATTTAAGGTTTAAAGACAGTACAGATC 146
 363 ----GTAATTTTCACTGTGTAAACAGTACAGATC 331

Search completed: February 8, 2005, 16:37:03
 Job time : 817.048 secs

RESULT 15
 BV097603/c 695 bp DNA linear STS 14-JAN-2004
 LOCUS RPAMMS00010113 Roche Palo Alto Mus musculus STS genomic, sequence
 DEFINITION
 tagged site.
 BV097603
 BV097603.1 GI:40810679
 STS
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 695)
 Unaka,J., Liao,G., Cheng,J., Nguyen,A., Bach,C., Puech,A.,
 McPherson,J.D., Fournier,D. and Feltz,G.
 Mus musculus SNPs
 Unpublished (2003)

TITLE
 JOURNAL
 COMMENT
 Contact: Jonathan Unaka
 Roche Palo Alto Genetics and Genomics Department
 Roche Palo Alto
 3431 Hillview Ave, Mailstop S3-1, Palo Alto, CA 94024, USA
 Tel: 6508555807
 Email: Jonathan.Unaka@roche.com
 Primer A: No primer submitted with this STS
 Primer B: No primer submitted with this STS.
 Location/Qualifiers

1..695
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="x-73613-72919-Al805974.7.1.95826"
 /clone_lib="Roche Palo Alto"
 /note="SNPs developed from assay sequences derived from 15
 different strains of mice (as of October 1, 2003). Those
 strains include A/J, A/HeJ, 129/Sv, AKR/J, B10.D2-H2/oshJ,
 BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J,
 NZB/MPJ, NZB/BlmJ, NZM/Lac, SPRET/Ei."
 <1..>695

ORIGIN

STS

Query Match 34.2%; Score 50; DB 11; Length 695;

Best Local Similarity 68.2%; Pred. No. 0.0032; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

Matches 107; Conservative 0; Mismatches 31; Indels 19; Gaps 2;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 193.675 Seconds
(without alignments)
3957.220 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146
Sequence: 1 gattctgttcgtcgtcctcc.....gtttaagacagctacaagatcc 146

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseq1980a:*
2: geneseq1990a:*
3: geneseq2000a:*
4: geneseq2001a:*
5: geneseq2001b:*
6: geneseq2002a:*
7: geneseq2002b:*
8: geneseq2003a:*
9: geneseq2003b:*
10: geneseq2003c:*
11: geneseq2003d:*
12: geneseq2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.2	35.8	259202	12 ADQ18492	Adq18492 Human sof
2	40.2	27.5	722	6 AAL4111	Aal41111 GAML rela
3	40.2	27.5	752	6 AAL4110	Aal41110 GAML rela
4	38	26.0	270	12 ADM80835	Adm80835 Human CAD
5	38	26.0	549	12 ADM80826	Adm80826 Human CAD
6	38	26.0	623	12 ADM80827	Adm80827 Human CAD
7	36	24.7	825	4 AAH03775	Aah03775 Human CDN
8	36	24.7	1626	4 AAH17575	Aah17575 Human CDN
9	36	24.7	2176	4 AAH17101	Aah17101 Human CDN
10	36	24.7	3037	12 ADQ24117	Adq24117 Human sof
11	34.6	23.7	556	10 ADB59026	Adb59026 Toxicity-
12	34.6	23.7	556	10 ADB53782	Adb53782 Primary r
13	34.4	23.6	149671	6 ABK84797	Abk84797 Human CDN
14	34.4	23.6	149671	9 ADB70361	Adb70361 Moesin cd
15	34.4	23.6	149671	12 ADJ37140	Adj37140 Human mal
16	34	23.3	65	6 ABN28389	Abn28389 Rat splic
17	34	23.3	476	3 AAZ50832	Aaz50832 Rat amelo
18	33.8	23.2	318	12 ADM80836	Adm80836 Human CAD
19	33.8	23.2	750	2 AAZ07020	Aaz07020 Amelogeni
20	33.8	23.2	793	2 AAZ07018	Aaz07018 Amelogeni
21	33.6	23.0	588	4 AAH10816	Aah10816 Human CDN

22	33.6	23.0	2497	4 AAH15986	Aah15986 Human CDN
23	32.6	22.3	5759	10 ADB52555	Adb52555 Caenorhab
24	32.6	22.3	28564	10 ADD46508	Add46508 Human gen
25	32.4	22.2	6081	6 ABL33429	Ab133429 Human imm
26	32.4	22.2	6081	6 ABN80193	Abn80193 Human che
27	32.4	22.2	15373	6 ABL32466	Ab132466 Human imm
28	32.2	22.1	802	2 AAZ07019	Aaz07019 Amelogeni
29	32.2	22.1	852	12 ADQ22977	Adq22977 Human sof
30	32	21.9	110000	12 ADN46845_17	Adn46845_17 o
31	32	21.9	110000	12 ADN47591_02	Adn47591_02 o
32	32	21.9	110000	12 ADN47591_03	Adn47591_03 o
33	32	21.9	110000	12 ADN46123_17	Adn46123_17 o
34	32	21.9	110000	12 ADN47209_02	Adn47209_02 o
35	32	21.9	110000	12 ADN47209_03	Adn47209_03 o
36	32	21.9	110000	12 ADN46464_17	Adn46464_17 o
37	32	21.9	110000	12 ADN47960_02	Adn47960_02 o
38	32	21.9	110000	12 ADN47960_03	Adn47960_03 o
39	31.8	21.8	300	2 AAZ14959	Aaz14959 Human gen
40	31.8	21.8	4350	8 ABQ83860	Abq83860 Human MDD
41	31.8	21.8	4410	8 ABQ83860	Abq83860 Human MDD
42	31.8	21.8	6032	6 AAS63310	Aas63310 Chemical1
43	31.8	21.8	16200	4 ABL13736	Ab113736 Drosophila
44	31.8	21.8	110000	2 AAT42063_14	Continuation (15 o
45	31.4	21.5	440	8 ABX50288	Abx50288 Bovine ES

ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
ID	ADQ18492	ADQ18492
AC	ADQ18492	ADQ18492
XX	26-AUG-2004	(first entry)
DT	26-AUG-2004	(first entry)
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
DE	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KW	Homo sapiens.	
OS	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002DUS-0429739P.	
XX	26-NOV-2002; 2002DUS-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PA	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
DR	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Example 2; SEQ ID NO 1311; 210bp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

CC can be applied on different species of goats. This polynucleotide
CC sequence represents a GAWL related X-chromosome DNA sequence of the
CC invention

XX SQ Sequence 752 BP, 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;

Query Match 27.5%; Score 40.2; DB 6; Length 752;
Best Local Similarity 93.3%; Pred. No. 0.055; 3; Indels 0; Gaps 0;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTCTCTGAGACAGCTTCAGTATGCTGT 45
DB 46 GATTTGTTTGCTGCTCTCTGAGACAGCTTCAGTATGCTGT 90

RESULT 4
ADM80835
ID ADM80835 standard; cDNA; 270 BP.

XX AC ADM80835;

XX DT 03-JUN-2004 (first entry)

XX DE Human CADECM-22 encoding cDNA SEQ ID NO:64.

XX KM human; cell adhesion and extracellular matrix protein; CADECM;
XX KM neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KM neurological disorder; developmental disorder;
XX KM connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KM Tangier disease; gene; ss.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT 36..212
XX FT /*tag= a
XX FT /product= "CADECM-22"

XX PN WO2004015396-A2.

XX PD 19-FEB-2004.

XX PF 12-AUG-2003; 2003WO-US025418.

XX PR 13-AUG-2002; 2002US-0403781P.

XX PR 30-AUG-2002; 2002US-0407034P.

XX PR 13-SEP-2002; 2002US-0410566P.

XX PR 24-SEP-2002; 2002US-0413482P.

XX PR 25-SEP-2002; 2002US-0413890P.

XX PR 08-NOV-2002; 2002US-0424904P.

XX PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX PA Eliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
XX PI Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;

XX DR MPI; 2004-191795/18.

XX P-PSDB; ADM80793.

XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 64; 272bp; English.

XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADECM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative

CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
CC cancer, obesity and Tangier disease.

XX SQ Sequence 270 BP, 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;

Query Match 26.0%; Score 38; DB 12; Length 270;
Best Local Similarity 75.8%; Pred. No. 0.19; 15; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GATTTGTTTGCTGCTCTCTGAGACAGCTTCAGTATGCTGTAGTAAATTCGAA 60
DB 47 GATTTATTTGCTGCTCTCTGAGACAGCTTCAGTATGCTGTAGTAAATTCGAA 106

QY 61 TT 62

DB 107 GT 108

RESULT 5
ADM80826
ID ADM80826 standard; cDNA; 549 BP.

XX AC ADM80826;

XX DT 03-JUN-2004 (first entry)

XX DE Human CADECM-13 encoding cDNA SEQ ID NO:55.

XX KM human; cell adhesion and extracellular matrix protein; CADECM;
XX KM neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KM neurological disorder; developmental disorder;
XX KM connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KM Tangier disease; gene; ss.
XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT 36..491
XX FT /*tag= a
XX FT /product= "CADECM-13"

XX PN WO2004015396-A2.

XX PD 19-FEB-2004.

XX PF 12-AUG-2003; 2003WO-US025418.

XX PR 13-AUG-2002; 2002US-0403781P.

XX PR 30-AUG-2002; 2002US-0407034P.

XX PR 13-SEP-2002; 2002US-0410566P.

XX PR 24-SEP-2002; 2002US-0413482P.

XX PR 25-SEP-2002; 2002US-0413890P.

XX PR 08-NOV-2002; 2002US-0424904P.

XX PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

XX PA Eliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
XX PI Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;

XX DR MPI; 2004-191795/18.

XX P-PSDB; ADM80784.

XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 55; 272bp; English.

XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,

complementary strand of a polynucleotide which comprises one of the 560

PS Claim 8; SEQ ID NO 16427; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 2176 BP; 683 A; 445 C; 394 G; 654 T; 0 U; 0 Other;

Query Match 24.7%; Score 36; DB 4; Length 2176;

Best Local Similarity 70.6%; Pred. No. 1.3;

Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTTGCTGCTCCTCGAGAGAGCCTTCAGTATGCTGTAGTAAATTTCATTT 63
DB 458 TTTCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACCTTT 399

QY 64 CCAATTTC 71
DB 398 CTCCTTCC 391

RESULT 10

ADQ24117/C

ID ADQ24117 standard; DNA; 3037 BP.

XX ADQ24117;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 6937.

XX soft tissue sarcoma; cytoskeletal; gene therapy; vaccine; screening; human; de.

OS Homo sapiens.

PN WO2004048938-A2.

PD 10-JUN-2004.

PF 26-NOV-2003; 2003WO-US038193.

PR 26-NOV-2002; 2002US-0429739P.

PA (PROT-) PROTEIN DESIGN LABS INC.

PI Aziz N, Ginsburg WM, Zlotnick A;

DR WPI; 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

PS Example 2; SEQ ID NO 6937; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated DNA of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

XX Sequence 3037 BP; 957 A; 596 C; 545 G; 919 T; 0 U; 20 Other;

Query Match 24.7%; Score 36; DB 12; Length 3037;

Best Local Similarity 70.6%; Pred. No. 1.4;

Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTTGCTGCTCCTCGAGAGAGCCTTCAGTATGCTGTAGTAAATTTCATTT 63
DB 913 TTTCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACTTTACCTTT 854

QY 64 CCAATTTC 71
DB 853 CTCCTTCC 846

RESULT 11

ADB59026

ID ADB59026 standard; DNA; 556 BP.

XX ADB59026;

DT 04-DEC-2003 (first entry)

DE Toxicity-related gene, SEQ ID 4052.

XX Toxic; toxin; gene expression profile; hepatotoxicity; liver; drug screening; toxicity assay; de.

OS Unidentified.

PN WO2003064624-A2.

PD 07-AUG-2003.

PF 31-JAN-2003; 2003WO-US003194.

PR 31-JAN-2002; 2002US-00060087.

PR 15-MAR-2002; 2002US-0364045P.

PR 15-MAR-2002; 2002US-0364055P.

PR 30-DEC-2002; 2002US-0436643P.

PA (GENE-) GENE LOGIC INC.

PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

DR WPI; 2003-689530/65.

PT Predicting a toxic effect of a compound, useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays, comprises preparing gene expression profile of tissue or cells exposed to the compound.

PS Claim 1; SEQ ID NO 4052; 1156pp; English.

XX The present invention relates to a method for predicting a toxic effect of a compound. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the

CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;

Query Match 23.7%; Score 34.6; DB 10; Length 556;
Best Local Similarity 88.1%; Pred. No. 2.3;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTGAGAGAGCTTCAGTATGCC 42
33 GATCTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCC 74

Db

RESULT 12
ADBS3782
ID ADBS3782 standard; DNA; 556 BP.

AC ADBS3782;
XX
DT 04-DEC-2003 (first entry)
XX

DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
XX
XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
XX WO2003065993-A2.
XX
XX 14-AUG-2003.
XX
PD 04-FEB-2003; 2003WO-US003482.
XX
PF 04-FEB-2003; 2003WO-US003482.
XX
XX 04-FEB-2003; 2002US-0353171P.
PR 13-MAR-2002; 2002US-036334P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 22-APR-2002; 2002US-0373602P.
PR 08-MAY-2002; 2002US-0374139P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394231P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442500P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elshoff M;
XX
XX MPI; 2003-731472/69.

XX
PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.

PS Claim 44; SEQ ID NO 4324; 874pp; English.

XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.

SO Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;

Query Match 23.7%; Score 34.6; DB 10; Length 556;
Best Local Similarity 88.1%; Pred. No. 2.3;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTCTGAGAGAGCTTCAGTATGCC 42
33 GATCTTGTTGCTGCTCTGAGAGAGCTTCAGTATGCC 74

Db

RESULT 13
ABK84797/c
ID ABK84797 standard; cDNA; 149671 BP.

AC ABK84797;
XX
DT 14-AUG-2002 (first entry)
XX

DE Human cDNA differentially expressed in granulocytic cells #1368.
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; Chromobits;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
PD 03-OCT-2001; 2001WO-US030821.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2001; 2000US-0237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI MPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.

PS Claim 1; SEQ ID NO 1368; 114bp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing the
 CC expression level to an expression level in an unactivated GC, where
 CC differential expression of Gs is indicative of GCA. Also included are
 CC modulating (M2) Gs by contacting GC with an agent that alters the
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent
 CC capable of modulating GCA or an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammation disease using the gene expression
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammation disease, by detecting the level of
 CC expression in a sample of the tissue of gene(s) from Gs, where the level
 CC of expression of the gene is indicative of inflammation; (4) treating
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease, by contacting a tissue having inflammation with an
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
 CC is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful
 CC for screening an agent capable of modulating GCA preferably in an
 CC inflammation in a tissue; M4 is useful for detecting an inflammation
 CC (especially chronic) in a tissue, an allergic response in a subject,
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.,
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
 CC disease, ulcerative colitis, periodontal disease, also bacterial
 CC infection, viral infection, parasitic infection, protozoal infection,
 CC fungal infection and M5 is useful for treating one of the above
 CC conditions. The present sequence represents a gene differentially
 CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 149671 BP; 45600 A; 33308 C; 32389 G; 38374 T; 0 U; 0 Other;

XX Query Match 23.6%; Score 34.4; DB 6; Length 149671;

XX Best Local Similarity 53.8%; Pred. No. 11;

XX Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

XX QY 3 TTTTGTTCCTGCTCTCTGAGAGACCTTCAGTATGCTGAGTAATAATTCCAATT 62

XX DB 33979 TTTATGTTCCTGAACCTGTACTCTCCACAGTCATGAGCTGAGTAATACACAACC 33920

XX QY 63 TCCAAATTCACAAGCTTGAATAATAAATCGCTCAGATTTTATTTAGGTTAAATT 122

XX DB 33919 TCCAGTTCTTAAAGCTAGTGTCTTAAATACTGAAGTCCTACTTTCTTAGGTGTATGGGG 33860

XX QY 123 AAGGTTTAAAGA 134

XX DB 33859 CAGGTTCCAGGA 33848

XX RESULT 14

XX ADB70361/C

XX ID ADB70361 standard; cDNA; 149671 BP.

XX ADB70361;

XX DT 04-DEC-2003 (first entry)

XX DE Moesin cDNA SEQ ID NO:53.

XX KW cancer; malignant pleural mesothelioma; MPW; lung adenocarcinoma;

XX KW squamous carcinoma; medulloblastoma; prostate cancer; breast cancer;

XX KW diffuse large B-cell lymphoma; follicular lymphoma; ovarian cancer;

XX KW human; gene; ss.

XX Homo sapiens.

XX MO2003021229-A2.

XX 13-MAR-2003.

XX 05-SEP-2002; 2002WO-US028203.

XX 05-SEP-2001; 2001US-0317389P.

XX 30-AUG-2002; 2002US-00236031.

XX (BGM) BRIGHAM & WOMEN'S HOSPITAL INC.

XX Gordon GJ, Jensen RV, Gullans SR, Bueno R;

XX MPI: 2003-290233/28.

XX P-PSDB; ADB70362.

XX Diagnosing cancer cells in tissue sample, or determining prognosis or

XX outcome of cancer patient, by calculating ratio of expression levels of

XX genes that are differentially expressed in cancer and non cancer tissues.

XX Claim 67; Page 181-263; 396pp; English.

XX The present invention describes a method (M1) for diagnosing the presence

XX of cancer cells or non-cancer cells in a tissue sample, or determining

XX the prognosis or outcome of a cancer patient. M1 involves providing a set

XX of genes that are differentially expressed in cancerous or non-cancerous

XX conditions, determining the expression levels of the set of genes and

XX calculating a ratio of the expression levels of the differentially

XX expressed genes. M1 is useful for diagnosing the presence of cancer cells

XX or non-cancer cells in a tissue sample, where the cancer is malignant

XX pleural mesothelioma (MPW), lung adenocarcinoma, squamous carcinoma,

XX medulloblastoma, prostate cancer, breast cancer, diffuse large B-cell

XX lymphoma, follicular lymphoma and ovarian cancer, and for determining

XX prognosis or outcome of a cancer patient. The ratio of expression levels

XX of differentially expressed genes is used as an indicator of cancer type,

XX cancer class, and/or cancer prognosis, all of which are useful for

XX determining a course of treatment of a patient. The present sequence

XX encodes a human protein which is used in an example from the present

XX invention.

XX Sequence 149671 BP; 45600 A; 33308 C; 32389 G; 38374 T; 0 U; 0 Other;

XX Query Match 23.6%; Score 34.4; DB 9; Length 149671;

XX Best Local Similarity 53.8%; Pred. No. 11;

XX Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

XX QY 3 TTTTGTTCCTGCTCTCTGAGAGACCTTCAGTATGCTGAGTAATAATTCCAATT 62

XX DB 33979 TTTATGTTCCTGAACCTGTACTCTCCACAGTCATGAGCTGAGTAATACACAACC 33920

XX QY 63 TCCAAATTCACAAGCTTGAATAATAAATCGCTCAGATTTTATTTAGGTTAAATT 122

XX DB 33919 TCCAGTTCTTAAAGCTAGTGTCTTAAATACTGAAGTCCTACTTTCTTAGGTGTATGGGG 33860

XX QY 123 AAGGTTTAAAGA 134

XX DB 33859 CAGGTTCCAGGA 33848

XX RESULT 15

XX ADJ37140/C

XX ID ADJ37140 standard; cDNA; 149671 BP.

XX ADJ37140;

XX 22-APR-2004 (first entry)

XX Human malignant pleural mesothelioma (MPW) cDNA #23.

XX Human; malignant pleural mesothelioma; MPW; gene; ss; tumour;

XX lung adenocarcinoma; squamous carcinoma; medulloblastoma;

XX prostate cancer; breast cancer; diffuse large B-cell lymphoma;

XX human; gene; ss.

XX Homo sapiens.

KW follicular lymphoma; ovarian cancer; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN US2003219760-A1.
 XX
 PD 27-NOV-2003.
 XX
 PF 05-SEP-2002; 2002US-00236031.
 XX
 PR 05-SEP-2001; 2001US-0317389P.
 PR 30-AUG-2002; 2002US-0407431P.
 XX
 PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 PI Gordon GJ, Jensen RV, Gullans SR, Bueno R;
 XX
 DR MPI; 2004-141744/14.
 DR P-PSDB; ADJ37141.
 XX
 PT Diagnosing the presence of cancer or non-cancer cells in tissue sample,
 PT useful for diagnosing malignant pleural mesothelioma comprises
 PT determining ratio of expression level of a set of genes expressed in
 PT cancer tissues.
 XX
 PS Claim 44; SEQ ID NO 53; 53pp; English.
 XX
 CC The invention relates to a method of diagnosing the presence of cancer
 CC cells or non-cancer cells in a tissue sample, determining prognosis or
 CC outcome of a cancer patient, selecting a course of treatment for a
 CC subject having or suspected of having malignant pleural mesothelioma
 CC (MPM) and evaluating treatment of MPM comprising determining the ratio of
 CC the expression level of a set of genes differentially expressed in a
 CC cancer tissue. The cancer is chosen from MPM, lung adenocarcinoma,
 CC squamous carcinoma, medulloblastoma, prostate cancer, breast cancer,
 CC diffuse large B-cell lymphoma, follicular lymphoma and ovarian cancer.
 CC The method is useful for diagnosing MPM in a subject suspected of having
 CC MPM which involves obtaining a tissue sample suspected of being cancerous
 CC from a subject and determining the expression of nucleic acid markers or
 CC its expression products in the tissue sample. This sequence represents
 CC human MPM cDNA of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification but was obtained in
 CC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 149671 BP; 45600 A; 33308 C; 32389 G; 38374 T; 0 U; 0 Other;
 Query Match 23.6%; Score 34.4; DB 12; Length 149671;
 Best Local Similarity 53.8%; Pred. No. 11;
 Matches 71; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
 QY 3 TTTTGTTCCTGCTGCTGCGGAGCAGCTTCAGTATGCTGTGAGTAAATTCCATT 62
 DB 33979 TTTATGTTCTTGAGACCTCTCACTCTCCACAGTCACTGAGCTAGATACACACACC 33920
 QY 63 TCCAAATTCACAGCTTGGAATAAATCTGCTCAGATTTTATTAGGTTAAATT 122
 DB 33919 TCCAGTTTCTTAGCTAGTGTCTTAAAACTGAAGTCTACTTTCTTAGGTGTATGGG 33860
 QY 123 AAGGTTTAAAG 134
 DB 33859 CAGGTTCCAGGA 33848

Search completed: February 8, 2005, 13:57:17
 Job time : 196.675 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 39.4319 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146
Sequence: 1 gattctgttgcctgcctcc.....gtttaagacagtcacagatc 146

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A COMB. seq.*
2: /cgn2_6/prodata/1/ina/5B COMB. seq.*
3: /cgn2_6/prodata/1/ina/6A COMB. seq.*
4: /cgn2_6/prodata/1/ina/6B COMB. seq.*
5: /cgn2_6/prodata/1/ina/PCTUS COMB. seq.*
6: /cgn2_6/prodata/1/ina/backfileseq.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	34	23.3	476	US-09-744-128-3
2	33.8	23.2	500	US-09-018-595B-3
3	33.8	23.2	500	US-09-324-709A-3
4	33.8	23.2	793	US-09-018-595B-1
5	33.8	23.2	793	US-09-324-709A-1
6	32.2	22.1	802	US-09-018-595B-2
7	32.2	22.1	802	US-09-324-709A-2
8	31.8	21.8	1830121	US-09-557-884-1
9	31.8	21.8	1830121	US-09-643-990A-1
10	31.8	21.8	1830121	US-09-329-960-1
11	31	21.2	174	US-09-513-999C-30345
12	30.8	21.1	1263	US-09-328-352-1155
13	30	20.5	20674	US-09-641-638-651
14	30	20.5	20674	US-10-170-097-651
15	30	20.5	1664976	US-08-916-421B-1
16	30	20.5	1664976	US-09-697-570-1
17	29.8	20.4	1673	US-09-221-017B-315
18	29.8	20.4	1317	US-09-710-279-3615
19	29.8	20.4	3149	US-09-710-279-358
20	29.8	20.4	4211	US-09-004-838-106
21	29.6	20.3	152331	US-09-128-155-16
22	29.6	20.3	176373	US-09-128-155-17
23	29.4	20.1	444	US-09-270-767-4451
24	29.4	20.1	444	US-09-270-767-19733
25	29.4	20.1	942	US-09-248-796A-6969
26	29.4	20.1	2136	US-08-946-475-8
27	29.4	20.1	2136	US-09-340-479-8

C 28	29.4	20.1	2698	3	US-08-946-475-1	Sequence 1, Appl
C 29	29.4	20.1	2698	4	US-09-340-479-1	Sequence 1, Appl
C 30	29.2	20.0	5039	4	US-09-386-816C-1	Sequence 1, Appl
C 31	29	19.9	2445	4	US-09-248-796A-6685	Sequence 6685, Ap
C 32	29	19.9	5159	2	US-08-146-930-3	Sequence 3, Appl
C 33	29	19.9	5159	3	US-08-146-930-3	Sequence 3, Appl
C 34	29	19.9	5159	5	PCT-US93-03993-3	Sequence 3, Appl
C 35	29	19.9	5385	4	US-09-920-804-1	Sequence 1, Appl
C 36	28.8	19.7	5201	4	US-09-540-236-182	Sequence 182, App
C 37	28.8	19.7	5063	1	US-08-185-432-1	Sequence 1, Appl
C 38	28.8	19.7	99629	4	US-09-596-002-37	Sequence 37, Appl
C 39	28.6	19.6	261	3	US-08-991-789A-19	Sequence 19, Appl
C 40	28.6	19.6	261	3	US-09-062-451-19	Sequence 19, Appl
C 41	28.6	19.6	261	4	US-09-598-326-19	Sequence 19, Appl
C 42	28.6	19.6	261	4	US-09-289-198-19	Sequence 19, Appl
C 43	28.6	19.6	261	4	US-09-429-755-19	Sequence 19, Appl
C 44	28.6	19.6	756	4	US-09-270-767-9938	Sequence 9938, Ap
C 45	28.6	19.6	864	4	US-09-248-796A-567	Sequence 567, App

ALIGNMENTS

```
RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27656/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3

Query Match          23.3%; Score 34; DB 4; Length 476;
Best Local Similarity 88.1%; Pred. No. 0.23;
Matches 37; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GATTGTGTTGCTGCTGCTGCTGAGACAGCTTCACTATGCC 42
    |||||
Db 59 GATCTGTTGCTGCTGCTGCTGAGACAGCTTGTCTATGCC 100

RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
```

```

;
; ZIP: 94304-1050
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 3.5 inch diskette
;
; COMPUTER: IBM compatible
;
; OPERATING SYSTEM: Microsoft Windows 95
;
; SOFTWARE: Wordperfect for windows 6.0,
;
; SOFTWARE: ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/018,595B
;
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: David J. Weitz
;
; REGISTRATION NUMBER: 38,362
;
; REFERENCE/DOCKET NUMBER: PELM-744
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (650) 493-9300
;
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 3:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 500 nucleotides
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: linear
;
US-09-018-595B-3

Query Match      23.2%; Score 33.8; DB 2; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.26;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 GATTTGTTGGCTGCTCTCTGGAGACAGCCTTCAGTATGCTGT 45
      80 GATTTATTGCTGCTGCTCTGGAGACAGCCTTTTGCATGCTCT 124

Db

RESULT 3
US-09-324-709A-3
;
; Sequence 3, Application US/09324709A
;
; Patent No. 6154707
;
; GENERAL INFORMATION:
;
; APPLICANT: Perkin-Elmer Corporation,
;
; APPLICANT: Applied Biosystems Division
;
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
;
; SITES
;
; NUMBER OF SEQUENCES: 11
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSER: David J. Weitz,
;
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
;
; STREET: 650 Page Mill Road
;
; CITY: Palo Alto
;
; STATE: California
;
; COUNTRY: USA
;
; ZIP: 94304-1050
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 3.5 inch diskette
;
; COMPUTER: IBM compatible
;
; OPERATING SYSTEM: Microsoft Windows 95
;
; SOFTWARE: Wordperfect for windows 6.0,
;
; SOFTWARE: ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/324,709A
;
; FILING DATE:
;
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: David J. Weitz
;
; REGISTRATION NUMBER: 38,362
;
; REFERENCE/DOCKET NUMBER: 16842-758
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (650) 493-9300
;
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 793 nucleotides
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-09-018-595B-1

Query Match      23.2%; Score 33.8; DB 2; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.31;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```

;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (650) 493-9300
;
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 3:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 500 nucleotides
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: linear
;
US-09-324-709A-3

Query Match      23.2%; Score 33.8; DB 3; Length 500;
Best Local Similarity 84.4%; Pred. No. 0.26;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 GATTTGTTGGCTGCTCTCTGGAGACAGCCTTCAGTATGCTGT 45
      80 GATTTATTGCTGCTGCTCTGGAGACAGCCTTTTGCATGCTCT 124

Db

RESULT 4
US-09-018-595B-1
;
; Sequence 1, Application US/09018595B
;
; Patent No. 5962233
;
; GENERAL INFORMATION:
;
; APPLICANT: Perkin-Elmer Corporation,
;
; APPLICANT: Applied Biosystems Division
;
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
;
; NUMBER OF SEQUENCES: 11
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSER: David J. Weitz,
;
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
;
; STREET: 650 Page Mill Road
;
; CITY: Palo Alto
;
; STATE: California
;
; COUNTRY: USA
;
; ZIP: 94304-1050
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 3.5 inch diskette
;
; COMPUTER: IBM compatible
;
; OPERATING SYSTEM: Microsoft Windows 95
;
; SOFTWARE: Wordperfect for windows 6.0,
;
; SOFTWARE: ASCII (DOS) TEXT format
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/018,595B
;
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: David J. Weitz
;
; REGISTRATION NUMBER: 38,362
;
; REFERENCE/DOCKET NUMBER: PELM-744
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (650) 493-9300
;
; TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 793 nucleotides
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
US-09-018-595B-1

Query Match      23.2%; Score 33.8; DB 2; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.31;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

RESULT 5
US-09-324-709A-1
Sequence 1, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-1
Query Match 23.2%; Score 33.8; DB 3; Length 793;
Best Local Similarity 84.4%; Pred. No. 0.31; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 7;
Cy 1 GATTGTGTCCTGCTGCTGAGAGAGAGCTTCAGTATGCTGT 45
Db 80 GATTGTGTCCTGCTGCTGAGAGAGAGCTTCAGTATGCTGT 124
RESULT 6
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 5962233
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2
Query Match 22.1%; Score 32.2; DB 2; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.95; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 8;
Cy 1 GATTGTGTCCTGCTGCTGAGAGAGAGCTTCAGTATGCTGT 45
Db 80 GATTGTGTCCTGCTGCTGAGAGAGAGCTTCAGTATGCTGT 124
RESULT 7
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362

REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match 22.1%; Score 32.2; DB 3; Length 802;
Best Local Similarity 82.2%; Pred. No. 0.95;
Matches 37; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GATTTGTTGCTGCTCTCTGGAGCAGCTTCAGTATGCTGT 45
Db 80 GATTTGTTGCTGCTCTCTGGAGCAGCTTCAGTATGCTCT 124

RESULT 8
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 21.8%; Score 31.8; DB 4; Length 1830121;
Best Local Similarity 57.6%; Pred. No. 14;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 46 GAGTAAATTTCCATTTCCATTTCCAGCTTGAATTAATTCGCTCAGATTTT 105
Db 1481806 GAGTAAATTTCCATTTTTTATCAAAAAGTATTAATTAATTAATTCAGCTTGAATTT 1481865

Qy 106 TATTTAGGTTAAATTTAAGGTTTAAAGACAGTACAGA 144
Db 1481866 TCCTTTGCACTAGTTTAAATAGTACAAAATTTCCAAA 1481904

RESULT 9
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Adams

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 21.8%; Score 31.8; DB 4; Length 1830121;
Best Local Similarity 57.6%; Pred. No. 14;
Matches 57; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 46 GAGTAAATTTCCATTTCCATTTCCAGCTTGAATTAATTCGCTCAGATTTT 105
Db 1481806 GAGTAAATTTCCATTTTTTATCAAAAAGTATTAATTAATTAATTCAGCTTGAATTT 1481865

Qy 106 TATTTAGGTTAAATTTAAGGTTTAAAGACAGTACAGA 144
Db 1481866 TCCTTTGCACTAGTTTAAATAGTACAAAATTTCCAAA 1481904

RESULT 10
US-10-329-960-1
Sequence 1, Application US/10329960

```
/ Patent No. 6742927
/ GENERAL INFORMATION:
/ APPLICANT: Fleischmann et al.
/ TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
/ Patent No. 6742927
/ TITLE OF INVENTION: Thereof, and Uses Thereof
/ FILE REFERENCE: P186P1
/ CURRENT APPLICATION NUMBER: US/10/329,960
/ CURRENT FILING DATE: 2003-01-02
/ PRIOR APPLICATION NUMBER: US 09/643,990
/ PRIOR FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: US 08/487,429
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: US 08/426,787
/ PRIOR FILING DATE: 1995-04-21
/ NUMBER OF SEQ ID NOS: 1
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 1830121
/ TYPE: DNA
/ ORGANISM: Haemophilus influenzae
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (4747)..(4747)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (9921)..(9921)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (10150)..(10150)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (29298)..(29298)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36543)..(36543)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36551)..(36551)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (36636)..(36636)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (40808)..(40810)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44416)..(44416)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44905)..(44905)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44975)..(44975)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45593)..(45593)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51309)..(51309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (10091)..(10091)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119924)..(119924)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
```

```

RESULT 13
US-09-641-638-651
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumentfeld, Marta
; APPLICANT: Bougenfeler, Lydie
; APPLICANT: Chumakov, Ilya

```



```
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5986..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
```

1	OTHER INFORMATION: 10-347-203	: polymorphic base A or G
2	NAME/KEY: allele	
3	LOCATION: 6464	
4	OTHER INFORMATION: 10-347-220	: polymorphic base A or G
5	NAME/KEY: allele	
6	LOCATION: 6534	
7	OTHER INFORMATION: 10-347-271	: polymorphic base A or T
8	NAME/KEY: allele	
9	LOCATION: 6611	
10	OTHER INFORMATION: 10-347-348	: polymorphic base A or G
11	NAME/KEY: allele	
12	LOCATION: 7668	
13	OTHER INFORMATION: 10-348-391	: polymorphic base A or G
14	NAME/KEY: allele	
15	LOCATION: 8608	
16	OTHER INFORMATION: 10-349-47	: polymorphic base C or T
17	NAME/KEY: allele	
18	LOCATION: 8658	
19	OTHER INFORMATION: 10-349-97	: polymorphic base A or G
20	NAME/KEY: allele	
21	LOCATION: 8703	
22	OTHER INFORMATION: 10-349-142	: polymorphic base G or C
23	NAME/KEY: allele	
24	LOCATION: 8777	
25	OTHER INFORMATION: 10-349-216	: deletion of CTG
26	NAME/KEY: allele	
27	LOCATION: 8785	
28	OTHER INFORMATION: 10-349-224	: polymorphic base G or T
29	NAME/KEY: allele	
30	LOCATION: 8926	
31	OTHER INFORMATION: 10-349-368	: polymorphic base C or T
32	NAME/KEY: allele	
33	LOCATION: 12171	
34	OTHER INFORMATION: 10-350-72	: polymorphic base C or T
35	NAME/KEY: allele	
36	LOCATION: 12429	
37	OTHER INFORMATION: 10-350-332	: polymorphic base C or T
38	NAME/KEY: allele	
39	LOCATION: 13341	
40	OTHER INFORMATION: 10-507-170	: polymorphic base A or G
41	NAME/KEY: allele	
42	LOCATION: 13492	
43	OTHER INFORMATION: 10-507-321	: polymorphic base A or C
44	NAME/KEY: allele	
45	LOCATION: 13524	
46	OTHER INFORMATION: 10-507-353	: polymorphic base C or T
47	NAME/KEY: allele	
48	LOCATION: 13535	

Query Match	20.5%	Score 30	DB 4	Length 20674
Best Local	59.3%	Pred. No. 13		
Matches 51	Conservative	0	Mismatches 35	Indels 0
			Gaps	0

QY	DB	QY	DB
46	11080	106	11140
GAGTAAATATTTCCATTTCCAAATTCACAAAGCTGGAAAATAAAATCTGCTCAGATTTT	GAAAAATTTTAAATATTTTAAATTTAATTAATAATTTAAATTTAAATTTT	TATTTAGGCTTAATTTAAGGCTTTA	AATTTATATTTAATTTAAATTTTAA
105	11139	131	11165

RESULT 14
US-10-170-097-651

Sequence 651, Application US/10170097
Patent No. 6794143
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueteler, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC CHROMATIDS INVOLVED IN ARACHIDONIC ACID METABOLISM
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM

```

1 FILE REFERENCE: GEN-T114X2D1
2 CURRENT APPLICATION NUMBER: US/10/170,097
3 CURRENT FILING DATE: 2002-06-10
4 PRIOR APPLICATION NUMBER: US 09/641,638
5 PRIOR FILING DATE: 2000-08-16
6 PRIOR APPLICATION NUMBER: US 09/502,330
7 PRIOR FILING DATE: 2000-02-11
8 PRIOR APPLICATION NUMBER: US 60/133,200
9 PRIOR FILING DATE: 1999-05-07
10 PRIOR APPLICATION NUMBER: US 09/275,267
11 PRIOR FILING DATE: 1999-03-23
12 PRIOR APPLICATION NUMBER: US 60/119,917
13 PRIOR FILING DATE: 1999-02-12
14 NUMBER OF SEQ ID NOS: 1304
15 SOFTWARE: Patent.pm
16 SEQ ID NO 651
17 LENGTH: 20674
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 FEATURE:
21 NAME/KEY: misc feature
22 LOCATION: 1123..3123
23 OTHER INFORMATION: 5' regulatory region
24 FEATURE:
25 NAME/KEY: exon
26 LOCATION: 3124..3297
27 OTHER INFORMATION: exon 1
28 FEATURE:
29 NAME/KEY: exon
30 LOCATION: 3871..4072
31 OTHER INFORMATION: exon 2
32 FEATURE:
33 NAME/KEY: exon
34 LOCATION: 5552..5633
35 OTHER INFORMATION: exon 3
36 FEATURE:
37 NAME/KEY: exon
38 LOCATION: 5758..5880
39 OTHER INFORMATION: exon 4
40 FEATURE:
41 NAME/KEY: exon
42 LOCATION: 5996..6099
43 OTHER INFORMATION: exon 5
44 FEATURE:
45 NAME/KEY: exon
46 LOCATION: 6349..6509
47 OTHER INFORMATION: exon 6
48 FEATURE:
49 NAME/KEY: exon
50 LOCATION: 7379..7522
51 OTHER INFORMATION: exon 7
52 FEATURE:
53 NAME/KEY: exon
54 LOCATION: 8645..8854
55 OTHER INFORMATION: exon 8
56 FEATURE:
57 NAME/KEY: exon
58 LOCATION: 12254..12340
59 OTHER INFORMATION: exon 9
60 FEATURE:
61 NAME/KEY: exon
62 LOCATION: 12854..13023
63 OTHER INFORMATION: exon 10
64 FEATURE:
65 NAME/KEY: exon
66 LOCATION: 13308..13429
67 OTHER INFORMATION: exon 11
68 FEATURE:
69 NAME/KEY: exon
70 LOCATION: 16567..16667
71 OTHER INFORMATION: exon 12
72 FEATURE:
73 NAME/KEY: exon

```

```

1 OTHER INFORMATION: exon 1
2 FEATURE:
3 NAME/KEY: exon
4 LOCATION: 3877..4072
5 OTHER INFORMATION: exon 2
6 FEATURE:
7 NAME/KEY: exon
8 LOCATION: 5552..5633
9 OTHER INFORMATION: exon 3
10 FEATURE:
11 NAME/KEY: exon
12 LOCATION: 5758..5880
13 OTHER INFORMATION: exon 4
14 FEATURE:
15 NAME/KEY: exon
16 LOCATION: 5996..6099
17 OTHER INFORMATION: exon 5
18 FEATURE:
19 NAME/KEY: exon
20 LOCATION: 6349..6509
21 OTHER INFORMATION: exon 6
22 FEATURE:
23 NAME/KEY: exon
24 LOCATION: 7379..7522
25 OTHER INFORMATION: exon 7
26 FEATURE:
27 NAME/KEY: exon
28 LOCATION: 8645..8854
29 OTHER INFORMATION: exon 8
30 FEATURE:
31 NAME/KEY: exon
32 LOCATION: 12324..12340
33 OTHER INFORMATION: exon 9
34 FEATURE:
35 NAME/KEY: exon
36 LOCATION: 12854..13023
37 OTHER INFORMATION: exon 10
38 FEATURE:
39 NAME/KEY: exon
40 LOCATION: 13308..13429
41 OTHER INFORMATION: exon 11
42 FEATURE:
43 NAME/KEY: exon
44 LOCATION: 16557..16667
45 OTHER INFORMATION: exon 12
46 FEATURE:
47 NAME/KEY: exon

```

LOCATION: 16775..16945
OTHER INFORMATION: exon 13
FEATURE:
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
FEATURE:
NAME/KEY: misc feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
FEATURE:
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
FEATURE:
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
FEATURE:
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
FEATURE:
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
FEATURE:
NAME/KEY: allele
LOCATION: 4088

OTHER INFORMATION: 12-206-366 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
Query Match 20.5%; Score 30; DB 4; Length 20674;
Best local similarity 59.3%; Pred. No. 13;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 46 GAGTAAATTTCCATTTCACAGCTTGAAATTAATAATCTGCCTCAGATTTT 105
DB 11080 GAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 11139
QY 106 TATTTAGGTTAAATTTAGCGTTA 131
DB 11140 AATTAAATTTAAATTTAAATTTTAA 11165
RESULT 15
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Buit et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22

```
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Mechanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
```

```
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664854)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

Query Match 20.54; Score 30; DB 4; Length 1664976;

Best Local Similarity 61.54; Pred. NO. 47;

Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
QY 68 TTTCACAGCTTGGAATAAAATCTGCTCAGATTTTATTAGGTTAAATTAAAGG 127
DB 1559577 TATGATGAGTTAGCATATATTTTAAAGACAGTTTATGAGATGTGATTGAAGG 1559636
QY 128 TTTAAGACAGTACAGAT 145
DB 1559637 GCTAAGATTTACTGAT 1559654
```

Search completed: February 8, 2005, 22:26:25
Job time : 49.543 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 196.735 Seconds

(without alignments)
4264.115 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146

Sequence: 1 gttctgttcgtcgtccctcc.....gtttaagacagacaaagac 146

Scoring table: IDENTITY_NNC

Gapop 10.0, Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues 8600550

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCF_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCFUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*

20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52.8	36.2	600	US-10-027-632-287050	Sequence 287050,
C 2	52.8	36.2	600	US-10-027-632-287050	Sequence 287050,
C 3	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 4	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 5	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 6	52.2	35.8	600	US-10-027-632-287049	Sequence 287049,
C 7	35	24.0	17	US-10-741-601-15294	Sequence 6937, A
C 8	35	24.0	17	US-10-741-601-15294	Sequence 6937, A
C 9	34.4	23.6	15	US-10-236-031B-53	Sequence 5660, Ap
C 10	34.4	23.6	15	US-10-236-031B-53	Sequence 5660, Ap
C 11	33.8	23.2	500	US-10-104-774-3	Sequence 1137, Ap
C 12	33.8	23.2	500	US-10-104-774-3	Sequence 1137, Ap
C 13	33.8	23.2	500	US-10-455-150-3	Sequence 3, Appli
C 14	33.8	23.2	500	US-10-455-150-3	Sequence 3, Appli

13	33.8	23.2	793	13	US-10-104-774-1	Sequence 1, Appli
14	33.8	23.2	793	16	US-10-455-150-1	Sequence 1, Appli
15	33.8	23.2	793	17	US-10-755-889-605	Sequence 605, App
16	33.8	23.2	38299	17	US-10-741-601-5755	Sequence 5755, Ap
C 17	33.8	23.0	101782	17	US-10-741-601-5661	Sequence 5661, Ap
C 18	33.6	23.0	43871	17	US-10-741-601-5628	Sequence 5628, Ap
C 19	33.6	23.0	84886	17	US-10-741-601-5721	Sequence 5721, Ap
C 20	32.6	22.3	5759	16	US-10-347-470A-7	Sequence 7, Appli
C 21	32.4	22.2	6081	15	US-10-311-455-1402	Sequence 1402, Ap
C 22	32.4	22.2	15373	15	US-10-311-455-439	Sequence 439, Appl
C 23	32.2	22.1	802	13	US-10-104-774-2	Sequence 2, Appli
C 24	32.2	22.1	802	16	US-10-455-150-2	Sequence 2, Appli
C 25	32.2	22.1	852	18	US-10-723-860-5797	Sequence 5797, Ap
C 26	32	21.9	1773	16	US-10-424-559-135058	Sequence 135058, Ap
C 27	32	21.9	289730	18	US-10-719-993-6780	Sequence 6780, Ap
C 28	31.8	21.8	3130	13	US-10-027-632-177165	Sequence 177165, Ap
C 29	31.8	21.8	3130	15	US-10-027-632-177165	Sequence 177165, Ap
C 30	31.8	21.8	4350	17	US-10-473-576-45	Sequence 45, Appl
C 31	31.8	21.8	4410	17	US-10-473-576-25	Sequence 25, Appl
C 32	31.8	21.8	6032	16	US-10-240-454-5	Sequence 5, Appli
C 33	31.8	21.8	1830121	14	US-10-328-960-1	Sequence 1, Appli
C 34	31.8	21.8	1830121	16	US-10-328-960-1	Sequence 1, Appli
C 35	31.8	21.8	1830121	18	US-10-158-865-1	Sequence 11205, A
C 36	31.6	21.6	1145	16	US-10-425-114-11205	Sequence 11205, A
C 37	31.6	21.6	1238	16	US-10-424-559-104040	Sequence 104040, A
C 38	31.6	21.6	561515	17	US-10-741-601-5682	Sequence 5682, Ap
C 39	31.4	21.5	440	9	US-09-983-965-217	Sequence 217, App
C 40	31.4	21.5	579	18	US-10-357-930-55270	Sequence 55270, A
C 41	31.2	21.4	650	13	US-10-027-632-226083	Sequence 226083, A
C 42	31.2	21.4	650	13	US-10-027-632-226084	Sequence 226084, A
C 43	31.2	21.4	650	15	US-10-027-632-226083	Sequence 226083, A
C 44	31.2	21.4	650	15	US-10-027-632-226083	Sequence 226083, A
C 45	31.2	21.4	672	16	US-10-424-559-113170	Sequence 113170, A

ALIGNMENTS

RESULT 1

US-10-027-632-287050/c

Sequence 287050, Application US/10027632

Publication No. US20020198371A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: Polymorphisms in the Human Genome

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 287050

LENGTH: 600

TYPE: DNA

ORGANISM: Human

US-10-027-632-287050

Query Match

Best Local Similarity

36.2%; Score 52.8; DB 13; Length 600;

75.9%; Pred. No. 5.9e-05;

	Matches	82; Conservative	1; Mismatches	13; Indels	12; Gaps	1;
Qy	1	GATTTGGTTGGCTCCTCTCGGGAGCAGCCCTTCACTTTCGCTGTGAGTAAAAATTC---	57			
Dd	245	GATTTATTTTGGCTCCTCTCGGGAGCAGCTTTTGCATGCTGTGAGTAAAAACACCCCT	186			
Qy	58	-----CAATTTCGAATTTCAAACTTCGAAATTTAAAAATTCGCC	96			
Dd	185	TGCATTAAGTCAGTGTCCAAATTTCAACAACTTGGACATTTAAAAATTCGCC	138			

RESULT 2
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
GENERAL INFORMATION

```

? TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
? TITLE OF INVENTION: Polymorphisms in the Human Genome
? FILE REFERENCE: 108827.129
? CURRENT APPLICATION NUMBER: US/10/027,632
? CURRENT FILING DATE: 2002-04-30
? PRIOR APPLICATION NUMBER: US 60/218,006
? PRIOR FILING DATE: 2000-07-12
? PRIOR APPLICATION NUMBER: US 60/198,676
? PRIOR FILING DATE: 2000-04-20
? PRIOR APPLICATION NUMBER: US 60/193,483
? PRIOR FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: US 60/185,218
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/167,363
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156,358
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146,002
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 325720
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 287050
? LENGTH: 600
? TYPE: DNA
? ORGANISM: Human
? US-10-027-632-287050

```

Query Match	36.2%	Score 52.8	DB 15	Length 600
Best Local Similarity	75.9%	Pred. No. 5	9e-05	
Matches	82	Conservative	1	Mismatches 13
				Indels 12
				Gaps 1

Qy	1	GATTTTGTTCGCTCCCTCCGAGACAGCCTTCAAGTACGCTGAGATTAATTTCT	---	57
Db	245	GATTTTATTTGCTGCTCTCTCTGAGACACTTTTGCCATGCGCTGTGAGTAAATTAACAC	CCT	186
Qy	58	-----CAATTTCAATTTCAACAAGTTGGAATTAATAATTTCTGCC		96
Db	185	TGCATTAAGTCAGTGTCCATTTTACAAACATTTGAGACATTAATAATTCGCT		138

```

RESULT 3
US-10-027-632--287049/C
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483

```

```

?
? PRIOR FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: US 60/185,218
?
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/167,363
?
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156,358
?
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146,002
?
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 325720
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 287049
?
? LENGTH: 600
?
? TYPE: DNA
? ORGANISM: Human
?
? US-10-027-632-287049

```

Query Match	35.8%	Score 52.2	DB 13	Length 600
Best Local Similarity	76.6%	Pred. No. 8	ae-05	
Matches 82	Conservative	0	Mismatches 13	Indels 12
				Gaps 1

QY 1 GATTTTGGTTGACCTGCTCTGGGAGCAGCCTTCAGTATGCTGTGAGTAAATTTTC --- 57
Db 245 GATTTTATTTGCTGCTCTCTGGGAGCAGCCTTTTGCATGCTCTGTGAGTAAATACACCCT 186

```
Qy      58 -----CAATTTCCAATTTCACAAGCTTGGAAATAAAAATCTGC   95
          ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     185 TGCATAAGTCAGTGTCCAAATTTCACAAAC TTGACATAAAAAATCTGC 139
```

RESULT 4
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9

```

1 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
2 FILE REFERENCE: 108827.129
3 CURRENT FILING DATE: 2002-04-30
4 PRIOR APPLICATION NUMBER: US 60/218,006
5 PRIOR FILING DATE: 2000-07-12
6 PRIOR APPLICATION NUMBER: US 60/198,676
7 PRIOR FILING DATE: 2000-04-20
8 PRIOR APPLICATION NUMBER: US 60/193,483
9 PRIOR FILING DATE: 2000-03-29
10 PRIOR APPLICATION NUMBER: US 60/185,218
11 PRIOR FILING DATE: 2000-02-24
12 PRIOR APPLICATION NUMBER: US 60/167,363
13 PRIOR FILING DATE: 1999-11-23
14 PRIOR APPLICATION NUMBER: US 60/156,358
15 PRIOR FILING DATE: 1999-09-28
16 PRIOR APPLICATION NUMBER: US 60/146,002
17 PRIOR FILING DATE: 1999-08-09
18 NUMBER OF SEQ. ID NOS: 325720
19 SOFTWARE: FastSeq for Windows Version 4.0
20 SEQ. ID NO 287049

```

```

; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049

```

Query Match	35.8%	Score 52.2	DB 15	Length 600
Seq. Similarity	76.6%	Pred. No. 8	8e-05	
Matches	82	Conservative	0	Mismatches 13
				Indels 12
				Gaps 1

QY	1	GATTTTGTTCCTCCTCTGGAGACAGCTTCAGTATCGCTGTGAGTAAATTTTC	57
DB	245	GATTTTATTTGGCTCCTCTCTGGAGACAGCTTTTGGCCATCGCTGTGAGTAAACACCCCT	186
QY	58	-----CAATTCGAATTCACAGCTTGGAAATAAATCTTCG	95

Db 185 TGCATAGTCAGTGTCCATTTCACAACTTGACATATAAATCTGC 139

RESULT 5
US-10-723-860-1311

/ Sequence 1311, Application US/10723860
/ Publication No. US20040253606A1
/ GENERAL INFORMATION:

/ APPLICANT: Aziz, Natsasha
/ APPLICANT: Ginsburg, Wendy M.
/ APPLICANT: Zlotnik, Albert
/ TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
/ FILE REFERENCE: 05882.0193.NPUS01
/ CURRENT APPLICATION NUMBER: US/10/723.860
/ PRIOR FILING DATE: 2003-11-26
/ PRIOR APPLICATION NUMBER: 60/429,739
/ NUMBER OF SEQ ID NOS: 8393
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 1311
/ LENGTH: 239202
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 35.8%; Score 52.2; DB 18; Length 259202;
Best Local Similarity 76.6%; Pred. No. 0.00092;
Matches 82; Conservative 0; Mismatches 13; Indels 12; Gaps 1;

QY 1 GATTTGTTGCTGCTCTGCGAGGAGCGCTTCAGTATGCTGTGATTAATTC--- 57

Db 68884 GATTTATTTGCTGCTCTGCGAGGAGCGCTTTTCCATCCTGTGTGTAACACCCCT 69943

QY 58 -----CAATTTCCATTTTCACAGCTTGGAATAAATCTGC 95

Db 69944 TGCATAGTCAGTGTCCATTTCACAACTTGACATATAAATCTGC 69990

RESULT 6
US-10-723-860-6937/C

/ Sequence 6937, Application US/10723860
/ Publication No. US20040253606A1
/ GENERAL INFORMATION:

/ APPLICANT: Aziz, Natsasha
/ APPLICANT: Ginsburg, Wendy M.
/ APPLICANT: Zlotnik, Albert
/ TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
/ FILE REFERENCE: 05882.0193.NPUS01
/ CURRENT APPLICATION NUMBER: US/10/723.860
/ PRIOR FILING DATE: 2003-11-26
/ PRIOR APPLICATION NUMBER: 60/429,739
/ NUMBER OF SEQ ID NOS: 8393
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6937
/ LENGTH: 3037
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1804)..(1823)
/ OTHER INFORMATION: n 18 a, c, g, or t
US-10-723-860-6937

Query Match 24.7%; Score 36; DB 18; Length 3037;
Best Local Similarity 70.6%; Pred. No. 6.5;
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTTGCTGCTCTGCGAGGAGCGCTTCAGTATGCTGTGATTAATTTCAATTT 63

Db 913 TTTCCTTGAAGTGTGAGAGTAATTCAGTTTCCGTATTAACATTACCTTT 854

QY 64 CCAATTTTC 71
Db 853 CTCCTTCC 846

RESULT 7
US-10-741-601-15294

/ Sequence 15294, Application US/10741601
/ Publication No. US20040166519A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001500
/ CURRENT APPLICATION NUMBER: US/10/741.601
/ PRIOR FILING DATE: 2003-12-22
/ NUMBER OF SEQ ID NOS: 26415
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15294
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-741-601-15294

Query Match 24.0%; Score 35; DB 17; Length 201;
Best Local Similarity 59.6%; Pred. No. 4.4;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 30 CCTTCAGTATGCTGTGAGTAAATTTCCAAATTTCCAGCTTGGAATTAATA 89

Db 15 CATTCTGATTTCCACTAGCAATACACAAAGTTCCAACTTCATATTTTGCAATTA 74

QY 90 ATCTGCTCAGATTTTATTTAGGTTAAATTTAAGGT 128

Db 75 TTATTTTCGTGTTTTTTTTTTTTTTAATATATATGT 113

RESULT 8
US-10-741-601-5660

/ Sequence 5660, Application US/10741601
/ Publication No. US20040166519A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001500
/ CURRENT APPLICATION NUMBER: US/10/741.601
/ PRIOR FILING DATE: 2003-12-22
/ NUMBER OF SEQ ID NOS: 26415
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5660
/ LENGTH: 189817
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)....(189817)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-741-601-5660

Query Match 24.0%; Score 35; DB 17; Length 189817;
Best Local Similarity 59.6%; Pred. No. 62;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 30 CCTTCAGTATGCTGTGAGTAAATTTCCAAATTTCCAGCTTGGAATTAATA 89

Db 118492 CATTCTGATTTCCACTAGCAATACACAAAGTTCCAACTTCATATTTTGCAATTA 118551

QY 90 ATCTGCTCAGATTTTATTTAGGTTAAATTTAAGGT 128

Db 118552 TTATTTTCGTGTTTTTTTTTTTTTTAATATATATGT 118590

Query Match	23.3%	Score	34	DB	10	Length	65
Best Local Similarity	88.1%	Pred.	No. 5.4				
Matches	37	Conservative	0	Mismatches	5	Indels	0
						Gaps	0

?
? Publication 3, Application US/10455150
? Publication No. US20040053102A1
? GENERAL INFORMATION:
? APPLICANT: Perkin-Elmer Corporation,
? TITLE OF INVENTION: Applied Biosystems Division
? SITES
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: David J. Weitz,
? Wilson Sonneti Goodrich & Rosati
? STREET: 650 Page Mill Road
? CITY: Palo Alto
? STATE: California
? COUNTRY: USA
? ZIP: 94304-1050
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 inch diskette
? COMPUTER: IBM compatible
? OPERATING SYSTEM: Microsoft Windows 95
? SOFTWARE: wordperfect for windows 6.0,
? ASCII (DOS) TEXT format
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/10/455,150
? FILING DATE: 04-Jun-2003
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/09/324,709A
? FILING DATE: 03-June-1999
? ATTORNEY/AGENT INFORMATION:
? NAME: David J. Weitz
? REGISTRATION NUMBER: 38,362
? REFERENCE/DOCKET NUMBER: 16842-758
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (650) 493-9300
?

```
TELEFAX: (650) 493-6811
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 500 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3

Query Match      23.2%; Score 33.8; DB 16; Length 500;
Best Local Similarity 84.4%; Pred. No. 14;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAAGTATGCTCT 45
DB 80 GATTTATTTGGCTGCTCTCTGGAGACGCTTTCATGCTCT 124

RESULT 13
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-1

Query Match      23.2%; Score 33.8; DB 13; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAAGTATGCTCT 45
DB 80 GATTTATTTGGCTGCTCTCTGGAGACGCTTTCATGCTCT 124

RESULT 14
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
;           Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
;           ASCII (DOS) TEXT format
```

```
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 793 nucleotides
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1

Query Match      23.2%; Score 33.8; DB 16; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAAGTATGCTCT 45
DB 80 GATTTATTTGGCTGCTCTCTGGAGACGCTTTCATGCTCT 124

RESULT 15
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 605
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-755-889-605

Query Match      23.2%; Score 33.8; DB 17; Length 793;
Best Local Similarity 84.4%; Pred. No. 16;
Matches 38; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GATTTGTTGGCTGCTCTCTGGAGACGCTTCAAGTATGCTCT 45
DB 80 GATTTATTTGGCTGCTCTCTGGAGACGCTTTCATGCTCT 124

Search completed: February 9, 2005, 06:51:15
Job time : 203.735 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: February 8, 2005, 12:58:25 ; Search time 1739 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-11

Perfect score: 146
Sequence: 1 gattctgcttcgctgcctcc.....gtttaagacagcacaagatc 146

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : Maximum Match 100%
Listing first 45 summaries

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	26.4	429	6	CB473334 sm68_E03.
2	37.4	25.6	633	6	CA682077 wlm24.pk0
3	36.2	24.8	1307	6	CG752765 P047-3-E0
C 4	36	24.7	331	5	BY496405 BY496405
5	36	24.7	441	1	AI798381 ct34c02.x
6	36	24.7	444	2	BE220076 hve6b03.x
7	36	24.7	448	1	AI338491 qg97h05.x
C 8	36	24.7	561	5	EX470683 DKF2P686N
9	36	24.7	586	2	BF001849 7994C09.x
10	36	24.7	588	2	BF048291 ncd17E04.
C 11	36	24.7	620	4	BG622884 602647718
C 12	36	24.7	707	4	CN360063 170005830
C 13	36	24.7	825	1	AU118731 AU118731
C 14	35.8	24.5	653	1	CE509949 ct1gr-988
15	35.8	24.5	799	9	CL504547 SALT_739-
16	35.6	24.4	430	7	W33932 mb54d02.x1
17	35.6	24.4	430	7	W40649 mc43h12.x1
18	35.6	24.4	500	2	BE519442 MA000433
19	35.6	24.4	536	7	W36345 mb72c12.r1
20	35.6	24.4	585	9	AY419441 mbs muscu
21	35.6	24.4	645	6	CD773419 AGENCOURT
22	35.6	24.4	707	6	CB056709 NISC_j119
23	35.6	24.4	812	3	AK029358 mbs muscu
24	35.6	24.4	843	6	CB588212 AGENCOURT

25	35.6	24.4	859	6	CB588525 AGENCOURT
26	35.6	24.4	865	6	CB589177 AGENCOURT
27	35.6	24.4	886	6	CB587051 AGENCOURT
28	35.6	24.4	889	6	CB589251 AGENCOURT
29	35.6	24.4	891	6	CB574837 AGENCOURT
30	35.6	24.4	913	6	CB590451 AGENCOURT
31	35.6	24.4	919	6	CB587332 AGENCOURT
32	35.6	24.4	928	6	CB590111 AGENCOURT
33	35.6	24.4	945	2	BB614068 AGENCOURT
34	35.6	24.4	990	7	W12906 ma8p03.r1
35	35.6	24.4	1006	7	W29475 mb99f11.r1
36	35.6	24.4	1020	7	W08102 mb40e08.r1
37	35.4	24.2	496	5	B074104 sag13f02
C 38	35.4	24.2	600	8	AQ428965 CTBT-B1-
39	35.2	24.1	689	8	B2017915 oen08g01
40	35	24.0	633	4	B1113076 602898855
C 41	35	24.0	872	8	BH167583 ENTRU14TR
42	35	24.0	893	8	AZ539059 ENTWC23TR
C 43	35	24.0	904	8	AZ692144 ENTUV53TR
C 44	34.8	23.8	478	8	AZ144107 SP_0009_B
C 45	34.8	23.8	965	9	CNS00302 AL064783 Drosophila

ALIGNMENTS

RESULT 1
CB473334/c 429 bp mRNA linear EST 26-MAR-2003
LOCUS sm68_E03.f en Sus scrofa cDNA 5', mRNA sequence.
DEFINITION CB473334
ACCESSION CB473334.1 GI:29279720
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
1 (bases 1 to 429)
Neilan, V.G., Kutish, G.F., Lu, Z., Zaak, A. and Rock, D.L.
Sequence analysis of African swine fever virus infected and non-infected porcine macrophage cDNA libraries
Unpublished (2003)

JOURNAL
COMMENT
Contact: Neilan JG
Plum Island Animal Disease Center
US Department of Agriculture, Agricultural Research Service
PO Box 848, Greenport, NY 11944-848, USA
Tel: 631 323 3133
Fax: 631 323 3044
Email: jneilan@pladc.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified by cross_match v0.990329 and Lucy v1.17p.
Seq primer: M13 Forward.
Location/Qualifiers

1..429
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/issue_type="lymphoid"
/cell_type="macrophage"
/lab_host="DH10B"
/clone_lib="en"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;
Library made from pools of polyA selected RNA, isolated at different times post-infection (0 to 16 hrs) from African swine fever virus (isolate Pretoriuskop/96/4) infected swine macrophages. Macrophages were derived from peripheral blood mononuclear cells cultured for 48 hrs on plastic in the presence of 30% U937 supernatant."

ORIGIN

Query Match 26.4%; Score 38.6; DB 6; Length 429;
Best local similarity 91.1%; Pred. No. 4.1;

Matches	41;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
Qy	1	GATTGTTGCTGCTGCTCTCTGGAGACACCTTCACTATGCTGT	45						
Db	357	GATTGTTGCTGCTGCTCTCTGGAGACACCTTCTTATGCTCT	313						
RESULT 2									
LOCUS	CA682077								
DEFINITION	wlm24.pk0030.h8 wlm24 Triticum aestivum cDNA clone wlm24.pk0030.h8								
ACCESSION	CA682077								
VERSION	CA682077.1								
KEYWORDS	EST.								
SOURCE	Triticum aestivum (bread wheat)								
ORGANISM	Triticum aestivum								
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.								
AUTHORS	1 (bases 1 to 633)								
TITLE	Tingey,S.V., Powell,W., Mollers,P., Dolan,M., Hailey,C., Yuan,Z.,								
JOURNAL	Miao,G., Carlier,N. and Hanafey,M.K.								
COMMENT	Dupont Wheat cDNA Sequence Unpublished (2002)								
FEATURES	Location/Qualifiers								
source	1..633								
	/organism="Triticum aestivum"								
	/mol_type="mRNA"								
	/culivar="Stephens"								
	/db_xref="taxon:4565"								
	/clone="wlm24.pk0030.h8"								
	/tissue_type="leaf"								
	/clone_lib="wlm24"								
	/note="Vector: pBluescript SK+, Site_1: EcoRI, Site_2: XhoI, wheat (Triticum aestivum L.) seedlings 24 hr after inoculation w/E. graminis"								
ORIGIN									
Query Match	25.6%; Score 37.4; DB 6; Length 633;								
Best Local Similarity	57.4%; Pred. No. 9.3;								
Matches	62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;								
Qy	29	GCCTTCAGTATGCTGCTGAGTAAATTTCCAAATTTCCAAATTTCAAGCTTGAAATMAA	88						
Db	510	GCCTTCCTAANAACGCTGATAAAAACCTGTACTCAGATCCANAGCGTGAAATCGA	569						
Qy	89	AATCGCTCAGATTTTATTATTAGGGTTAAATTTAAGGGTTTAGACA	136						
Db	570	AATCCCANNAAGCGTTGCTCTNNAAGATCNAATTCAGAGATTTCACA	617						
RESULT 3									
LOCUS	CG752765								
DEFINITION	P047-3-B05.ya Ppa BCORI BMC Library Pristionchus pacificus genomic								
ACCESSION	CG752765								
VERSION	CG752765.1								
KEYWORDS	GSS.								
SOURCE	Pristionchus pacificus								
ORGANISM	Pristionchus pacificus								
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Nematoda; Diplogasteridae; Pristionchus.								
	1 (bases 1 to 1307)								

AUTHORS Striñivasan,V., Sinz,W., Jesse,T., Wälgert-Berbolte,L., Jansen,K.,
 Buitfer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode *Pristionchus*
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel.: 00497071601371
 Fax: 00497071601498
 Email: ralf.sommer@tuebingen.mpg.de
FEATURES Class: BAC ends.
 location/Qualifiers
 1..1307
 /organism="Pristionchus pacificus"
 /mol_type="genomic DNA"
 /strain="California"
 /db_xref="taxon:54126"
 /clone_lib="Ppa ECoRI BAC Library"
 /note="The library was generated by a partial digest of
 the genomic DNA with ECoRI and cloning into the BAC
 vector."

Query Match	24.8%	Score 36.2	DB 9	Length 1307
Best Local Similarity	62.9%	Pred. No. 22		
Matches	56	Conservative 0	Mismatches 33	Indels 0
Qy	52	AAATTCACATTTCCAAATTCACAGCTTGGAATAAATCTGCTCAGATTTTATTTA	111	
Db	1162	AAATTCACAAATCCAAATTCCTTAATCCCTCGAAATATAATTTTATTTATTTTGTGTTTT	1223	
Qy	112	GGGTTAAATTTAAGGTTTAAGACAGTAC	140	
Db	1222	GGGAAAAAAAAAAAAAAAAAATAATAAAC	1250	
RESULT 4				
BY496405/c				
LOCUS				
DEFINITION	BY496405 RIKEN full-length enriched, bone marrow macrophage Mus	331 bp	linear	EST 14-DEC-2002
ACCESSION	musculus cDNA clone I83045E16.3, mRNA sequence.			
VERSION	BY496405			
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
REFERENCES	1 (bases 1 to 331)			
REFERENCES	Ozaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamanka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Haegawa, Y., Nogami, A., Schombach, C., Gotohori, T., Balderelli, R., Hill, D. P., Bult, C., Home, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brack, D., Brusil, V., Chotia, C., Cobani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. P., Forrest, A., Fraser, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustrich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagawa, A., Kurochkin, I. V., Lee, Y., Lennarz, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sardinia, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Vervaeke, R., Wagner, C., Weinsteiner, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyman-Boris, A., Yanagisawa, M., Yang, I.,			

Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Saito, K., Sato, K., Shiraki, T., Waki, K., Kawai, J., Alizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kogawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasuniishi, A., Yoshino, M., Materon, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL MEDLINE PUBLISHED
22354683
12466851

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/
Alizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN Integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source Location/Qualifiers
1..331
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1830045E16"
/issue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow macrophage"

ORIGIN

Query Match 24.7%; Score 36; DB 5; Length 331;
Best Local Similarity 58.3%; Pred. No. 21;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 38 ATGCTGTGAGTAAATTTCCATTTTCACAGCTTGAATAAATTCGCT 97
|||||
Db 154 ATGTTGGAGCTCAATGCGCATTAAGATTTTAACCAACAGTGGAGAGACTT 95
|||||

QY 98 CAGATTTTATTTAGGTTAAATTTAAGGTTTAAGCAGTACAGAT 145
|||||
Db 94 CAAATTTTATTTCTATTAGTTTAAATTTAATTAATGAATATAT 47
|||||

RESULT 5
LOCUS A1798381 441 bp mRNA linear EST 06-JUL-1999
DEFINITION tr34c02.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2220194 3', mRNA sequence.
ACCESSION A1798381
VERSION A1798381.1 GI:5363853
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 441)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Christopher Miskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: www.bio.lnl.gov/bdrp/image/image.html
Seq primer: -40UP from Glbco
High quality sequence stop: 395.

FEATURES

source Location/Qualifiers
1..441
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2220194"
/issue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Ov23"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors; metastasis positive. Life Technologies catalog #: 11534-013"

ORIGIN

Query Match 24.7%; Score 36; DB 1; Length 441;
Best Local Similarity 70.6%; Pred. No. 22;
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTTGTTGCTGCTCTCTGAGCAGCCTTCACTATGCTGTAGTAATTTCCATTT 63
|||||
Db 69 TTTCCTTGAAGCTGCTCATGAGACTTAATTCAGTTTCTGTATTAACTTACACTT 128
|||||

QY 64 CCAATTTC 71
|||||
Db 129 CTCCTTCC 136
|||||

ORIGIN

RESULT 6
LOCUS BE220076 444 bp mRNA linear EST 03-JUL-2000
DEFINITION hv66h03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3178421 3', mRNA sequence.
ACCESSION BE220076
VERSION BE220076.1 GI:8907394
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 444)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.

Contact: Robert Straubergs, Ph.D.
Email: cgabps-r@mail.nih.gov
Tissue Procurement: Christopher Moskalk, M.D., Ph.D., Michael R. Emmert-buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: infoimage@lntl.gov
Seq primer: -40up from Gbbco
High quality sequence stop: 388.

FEATURES	Location/Qualifiers
source	1. .444

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3178421"
/tissue_type="carcinoid"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lu24"
/note="Organ: Lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were
made in vitro. Following HAP purification, this DNA was
used as tracer in a subtractive hybridization reaction.
The driver was PCR-amplified cDNAs from a pool of 5,000
clones made from the same library (clonides
141920-1417991 and 150904-1522439). Subtraction by Bento
Soares and M.Fátima Bonaldo."

```

ORIGIN

Query Match	24.7%	Score 36	DB 2	Length 444
Best Local Similarity	70.6%	Prod. NC	22	
Matches 48, Conservative	0	Mismatches	20	Indels 0
				Gaps 0

Dy 4 TTTCCTTGCGCCCTCCGTGGAGCAGCCTTACGATGCCTGTGAGTAATAATTCCAACTTT 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 TTTCCTTGACTGCCCATGAGAGCAATTAATTCAGTTTTCTGTATTACAATTACACTTT 124

Qy	64	CCAATTTC	71
Db	125	CTCCTTCC	132

	RESULT	7
A1338491		
LOCUS		
DEFINITION		
COSSESTON		
A1338491	448 bp	mRNA linear EST J3-FEB-1996
cg97n05.x1 Soares_total_fetus NbZHF8_9w Homo sapiens CDNA clone		
IMAGE:1939353 3', mRNA sequence.		
M2206A01		

ACCESSION	A1338491	
VERSION	A1338491.1	GI:4075418
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE
1 (pages 1 to 448)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.

AUTHORS NCIC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP)

Tumor Gene Index
JOURNAL
Unpublished (1997)

COMMENT **Contact:** Robert Strausberg, Ph.D.

Email: c9abbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1345 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 447.

FEATURES
Source

Location/Qualifiers
1. .448

```

/organism="Homo sapien
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1939353"
/dev_stage="8-9 weeks"

```

"/clone_11fb=" Soares total fetus Nb2H9b_9w"
/note="Vector: pT7T3- β -lac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand
was prepared from mRNA obtained from pooled 8-9 week
total fetal tissue material with a Not I - oligo(dT) primer [5']
TGTCACGATCGAGGAGGAGCGCGCGCTTATTTTATTTTATTTT 3']
(Pharmacia), digested with Not I and Eco RI adaptors
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match	24.7%;	Score 36;	DB 1;	Length 448;
Best Local Similarity	70.6%;	Pred. No. 22;		
Matches 48;	Conservative 0;	Mismatches 20;	Indels 0;	Gaps 0;

4 ATTGTTTGCTGCTCCTCGGAGCAGCCTTCAATGCTGTGATAAATTTCCAATT 63
 69 TTTTCCTGACTGCTCATGAGACTAAATTCAGTTTTCCTGTATTACATTTACACTTT 128

QY	64	CCAATTTC	71
Db	129	CTCCTTCC	136

RESULT 8	LOCUS	DEFINITION	ACCESION
EX470683/c	EX470683	561 bp mRNA linear EST 04-SEP-2003 DXYZP686N21110.r1 686 (synonym: h1cc3) Homo sapiens cDNA clone DXYZP686N21110.5', mRNA sequence. BX470683	U00000

ACCESSION	DA7/0883	GI:31665009
VERSION	BX470683.1	
KEYWORDS	EST.	

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 561)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS Koehler, K., Beyer, A., Mewes, H. W., Well, B., Amid, C., Osanger, A., Pöb, G., Han, M. and Wiemann, S.
TITLE EST (Koehler, K., Beyer, A., Mewes, H. W., Well, B., Amid, C., et al.)
JOURNAL Unpublished (2003)
COMPASS COMPASS - WING

COMMENT **Contact: MIPS**

FEATURES

Ingolstaedter Landstr. 1, D-85764 Neuberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMPZ (Biomedical Research Center at the Heinrich-
Heine-University, Duesseldorf/Germany) within the cDNA sequencing
consortium of the German Genome Project. No 3' sequence available
This clone (DKFZp686N21110) is available at the RZPD in Berlin.
Please contact the RZPD: Resourcenentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/d_xref="taxon:9606"
/clone="DKFZp686n2110"
/dev_stage="adult"
/lab_host="DH10B"
/node_1b="c86 (synonym: hicc3)"
note=Vector: pTribEX2, site_1: sfla, site_2: sflb;

```


ORIGIN CDNA-collection"

Query Match 24.7%; Score 36; DB 5; Length 561;
 Best Local Similarity 70.6%; Pred. No. 22;
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTCAAGTATGCTGTAGTAAATTTCCAAATTT 63
 DB 211 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAAACATTTACACTTT 152

QY 64 CCAATTTTC 71
 DB 151 CTCCTTCC 144

RESULT 9 BF001849 586 bp mRNA linear EST 06-OCT-2000
 LOCUS 7994609.x1 NCI_CGAP_Col6 Homo sapiens CDNA clone IMAGE:3314128 3',
 DEFINITION mRNA sequence.

ACCESSION BF001849 GI:10702124
 VERSION BF001849
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 586)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Sequencing by: Greg Lennon, Ph.D.

CDNA distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL, send email to:

infoimage.lnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 471.

FEATURES Location/Qualifiers

1..586

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3314128"

/tissue_type="colon tumor, RER+"

/lab_host="DH10B"

/clone_1b="NCI CGAP Col6"

/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Site 1: Not 1; Site 2: Eco RI;

Plasmid DNA from the normalized library NCI_CGAP_Col6 was

prepared, and 88 circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneids 1057416-1061255, and 1144584-1145351).

Subtraction by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 24.7%; Score 36; DB 2; Length 586;
 Best Local Similarity 70.6%; Pred. No. 22;
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTCAAGTATGCTGTAGTAAATTTCCAAATTT 63
 DB 65 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAAACATTTACACTTT 124

QY 64 CCAATTTTC 71
 DB 125 CTCCTTCC 132

RESULT 10 BF448291 588 bp mRNA linear EST 01-DEC-2000
 LOCUS nad17f04.x1 NCI_CGAP_Lu24 Homo sapiens CDNA clone IMAGE:3365887 3',
 DEFINITION mRNA sequence.

ACCESSION BF448291 GI:11513814
 VERSION BF448291
 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 588)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Sequencing by: Greg Lennon, Ph.D.
 CDNA distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL, send email to:

infoimage.lnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 465.

FEATURES Location/Qualifiers

1..588

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3365887"

/tissue_type="carcinoid"

/lab_host="DH10B"

/clone_1b="NCI CGAP Lu24"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Plasmid DNA from the normalized

library NCI CGAP Lu5 was prepared, and 88 circles were

made in vitro. Following HAP purification, this DNA was

used as tracer in a subtractive hybridization reaction.

The driver was PCR-amplified cDNAs from a pool of 5,000

clones made from the same library (cloneids

1414920-1417991 and 1520904-1522439). Subtraction by Bento

Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 24.7%; Score 36; DB 2; Length 588;
 Best Local Similarity 70.6%; Pred. No. 22;
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 4 TTGTGTTGCTGCTCTGAGAGAGCCTTCAAGTATGCTGTAGTAAATTTCCAAATTT 63
 DB 69 TTTCCTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAAACATTTACACTTT 128

QY 64 CCAATTTTC 71
 DB 129 CTCCTTCC 136

RESULT 11 BG622884/c 620 bp mRNA linear EST 18-APR-2001
 LOCUS 602647718F1 NIH_MGC_79 Homo sapiens CDNA clone IMAGE:4769213 5',
 DEFINITION mRNA sequence.
 ACCESSION BG622884

VERSION BG622884.1 GI:13674255
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 620)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov
 Plate: L1CM1634 row: C column: 06
 High quality sequence stop: 618.
 Location/Qualifiers
 1. 620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4769213"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_1ib="NIH MGC 79"
 /note="Organ: Placenta; Vector: PDNR-LIB (Clontech); site_1: stit (ggcgctggcc); site_2: stit (ggcgctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 24.7%; Score 36; DB 4; Length 620;
 Best Local Similarity 70.6%; Pred. No. 23;
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 QY 4 TTTGTTTGCTGCTGCTCGGAGCAGCCTTCAGTATGCTGTGAGTAAATTTTCAATT 63
 DB 513 TTTCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACATTACACTTT 454
 QY 64 CCAATTTC 71
 DB 453 CTCCTTCC 446

RESULT 12
 CN360063/c 707 bp mRNA linear EST 16-MAY-2004
 LOCUS 17000583090269 GRN_PREHEP Homo sapiens CDNA 5', mRNA sequence.
 ACCESSION CN360063
 VERSION CN360063.1 GI:47359997
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 707)
 Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R., Lebkowski, J. and Stanton, L.W.
 Transcriptional characterization elucidates signaling networks that control human ES cell growth and differentiation
 Nat. Biotechnol. 22 (6), 707-716 (2004)

COMMENT Contact: Brandenberger R
 Regenerative Medicine
 Genon Corporation
 230 Constitution Drive, Menlo Park, CA 94025, USA
 Tel: 650 473 8658
 Fax: 650 473 7760
 Email: rbrandenberger@genon.com
 Insert length: 707 Std Error: 0.00.
 Location/Qualifiers
 1. 707
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cells, DMSO-treated H9 cell line"
 /clone_1ib="GRN_PREHEP"
 /note="oligo dT primed, full-length enriched cDNA library from DMSO-treated hES cell line H9 (p22) maintained in feeder-free conditions"

ORIGIN
 Query Match 24.7%; Score 36; DB 7; Length 707;
 Best Local Similarity 70.6%; Pred. No. 23;
 Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 QY 4 TTTGTTTGCTGCTGCTCGGAGCAGCCTTCAGTATGCTGTGAGTAAATTTCAATT 63
 DB 596 TTTCTTGACTGCTCATGAGAGCTAAATTCAGTTTCTGTATTAACATTACACTTT 537
 QY 64 CCAATTTC 71
 DB 536 CTCCTTCC 529

RESULT 13
 AU118731/c 825 bp mRNA linear EST 01-AUG-2002
 LOCUS AU118731 HEMBA1 Homo sapiens CDNA clone HEMBA1004241 5', mRNA
 DEFINITION sequence.
 ACCESSION AU118731
 VERSION AU118731.1 GI:10933870
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 825)
 Oca, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 229-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute
 Location/Qualifiers
 1. 825
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HEMBA1004241"
 /tissue_type="whole embryo, mainly head"
 /dev_stage="embryo, 10 weeks"
 /clone_1ib="HEMBA1"
 /note="Vector: PME18SFL3"

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 2743.23 Seconds
(without alignments)
8481.439 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492
Sequence: 1 atggggagactgattgttgc.....aactctctccgcgcctcca 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Genbank: 1: gb_ba: 2: gb_ncg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pac: 7: gb_ph: 8: gb_dl: 9: gb_pr: 10: gb_ro: 11: gb_ste: 12: gb_sy: 13: gb_un: 14: gb_vt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	339	66.9	5712	4	AB091793 Equus cab
2	303.2	61.6	6451	4	AB091789 Bos tauru
3	302.8	61.5	7425	4	AB091791 Sus scrof
4	274.6	55.8	8810	2	AY040206 Homo sapi
5	274.6	55.8	158142	2	AL357130 Homo sapi
6	274.6	55.8	259202	9	AC002366 Human Xp2
7	273	55.5	6465	9	AB091781 Pan trogl
8	261	53.0	5684	9	AB091785 Lemur cal
9	256.6	52.2	6442	9	AB091783 Saimiri s
10	248.8	50.6	5591	4	AB091794 Equus cab
11	245.2	49.8	5562	9	AB091787 Ocolemur
12	219.6	44.6	1935	9	D83730 Homo sapien
13	166.2	33.8	6264	4	AB091790 Bos tauru
14	160.6	32.6	9384	10	AF294397 Mus muscu
15	160.6	32.6	95826	10	AL805974 Mus muscu
16	150	30.5	695	11	BV089295 RPPAMSE00
17	150	30.5	695	11	BV097603 RPPAMSE00
18	142.2	28.9	212866	2	AC093946 Rattus norv
19	142.2	28.9	234471	2	AC121424 Rattus norv

20	127.6	25.9	8004	9	AB091786 Lemur cat
21	116.2	23.6	7454	9	AB091784 Saimiri s
22	96.8	19.7	463	10	D83063S2
23	95.4	19.4	177654	2	AP000918
24	95.4	19.4	200214	9	AC013412 Homo sapi
25	94.8	19.3	7163	9	AB091782 Pan trogl
26	94.8	19.3	38765	9	BS000568 Pan trogl
27	94.8	19.3	177726	9	AC145770 Pan trogl
28	94.8	19.3	190089	9	BS000576 Pan trogl
29	93	18.9	6931	4	AB091792 Sus scrof
30	82.8	16.8	2501	9	D83729
31	73	14.8	363	11	BV089294 RPPAMSE00
32	73	14.8	363	11	BV097602 RPPAMSE00
33	52.8	10.7	226	10	S74899
34	52.8	10.7	727	10	MUSMAMEB
35	52.8	10.7	765	4	AB032194
36	52.8	10.7	789	4	AB032193
37	52.8	10.7	798	10	BC059090
38	52.8	10.7	799	10	MUSMAMEA
39	51.6	10.5	166426	9	AC025505
40	51.6	10.5	187161	9	AC030824
41	51.2	10.4	399	10	RR07054
42	51.2	10.4	457	10	RNU60564
43	51.2	10.4	476	6	AR452534
44	51.2	10.4	753	10	RNU60562
45	51.2	10.4	780	10	U01245

ALIGNMENTS

RESULT 1
AB091793
LOCUS
DEFINITION
AB091793
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AB091793
Equus caballus AMELX gene for amelogenin, partial cds.
AB091793.1 GI:29126038

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
1 Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

FEATURES
source
1. 5712
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/sex="male"
174. .5712
/gene="AMELX"
join(174. .233,1578. .1589)
/gene="AMELX"
join(1590. .1643,3566. .3613,4993. .5037,5305. .55712)
/gene="AMELX"
/product="amelogenin"
/protein_id="BAC6112.1"
/db_xref="GI:29126039"

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
TITLE
JOURNAL

Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The ameglinin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
12672962
22608569
2 (bases 1 to 7425)
Iwase, M., Saita, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Soken-dai), Department of Biosystems Science;
Shonan Kokuai-inura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
Location/Qualifiers
1..7425
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/sex="male"
1491..7425
/gene="AMELX"
/join(1491..1548,3140..3151)
/gene="AMELX"
/join(3152..3205,5316..5363,6619..6660,6717..6761,
7033..7425)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC6110.1"
/db_xref="GI:29126035"
/translation="MGTWILFACLLGAAPMP,PPHGHGYINFSYEDLYLEAIRID
RRAFLTPKLYOMIRHPTYSYGYRPMGKMLHQIIPVVSQOTPOSHLQPHNHPM
VPAQDFIQQGPMPLPGHSMPTQHNQNLPLPAQDFPQVQPPQPPHQPQPS
MHPQLPLPPLPMPMSQSLPLDPLLEAMPXT"

ORIGIN

Query Match 61.5%; Score 302.8; DB 4; Length 7425;
Best Local Similarity 79.8%; Pred. No. 2.5e-75;
Matches 399; Conservative 0; Mismatches 87; Indels 14; Gaps 3;

QY 1 ATGGGACCTGATTTTGTTCCTGCTGCTGAGACAGCTTGTCTATGCCGTGAGT 60
DB 3152 ATGGGACCTGATTTTGTTCCTGCTGCTGAGACAGCTTGTCTATGCCGTGAGT 3211
QY 61 AAAATACCCC-----TGCATATATTCAATTTCACAGCTTGAAATAAAGTGGCC 114
DB 3212 AAAACACCCCTTCGATACGATGTCATATGTCATACGCTTGAAATAAATACCAACC 3271
QY 115 CACAGTTGTTAACTTTAGGGTTTAAAGACGTAACAAGATAGATGTCCTCAATGCTCT 174
DB 3272 CCGAGTTGATACGTTAGGGATGAAACAGTAGAATACATTTCTCTAAATGACTCA 3331
QY 175 GTGTTTAAAGAACCTTGAAAGAGCTTGTATATAAAAAATATATATCCAGATGCTCC 234
DB 3332 GTGTTTAAAGAACCTTGAAAGAGCTTGTATATAAAAAATATATATCCAGATGCTCT 3391
QY 235 ACCCAAGA--CTGATTCAGTAGACAGAGAGTGGGGGAGTGGCCAGAGACTTGCATTTTA 292
DB 3392 GCCAAAGATCTGATTCAGTAGAGCTGGGGTGG-----GCCCATGACTCTCATTTTA 3445
QY 293 ACAAGACCTCAGAGATCTGAGAGCAATTAATCTGTAATATCATCCCATCTCT 352
DB 3446 ACAAGACCTCAGAGAGCTCTTGAACAATTAAGCTTGAATATCATCATCTGTCTCT 3505
QY 353 AGATGAGAGAACTTTAGAAAGGACCTTGAAAGGCTTCAGAGAAAGTGTCTGAACAG 412
DB 3506 AGAAGTGGGAATTTTGAAGTGAACCTTTGAGAGGCTTCAGAGAAATGCTGAAGAGC 3565

QY 413 CTTAGGCAATACTACAAAATGCCAATTTTCTTAAACCAATTTCTACAGATGTC 472
DB 3566 CTTAGGCAATACTACAAAATGCCAATTTTCTTAAATTCGCAAGTGTCC 3625
QY 473 AACTCTCTCTGCTCCCTCCA 492
DB 3626 AATATCTCTCTCTCTCCA 3645

RESULT 4
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AY040206
AY040206
Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
AY040206.1 GI:15028582
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartini; Homidae; Homo.
Hart, P.S., Hart, T.C., Simmer, J.P. and Wright, J.T.
A nomenclature for X-linked amelogenesis imperfecta
Arch. Oral Biol. 47 (4), 255-260 (2002)
21920287
11922868
2 (bases 1 to 8810)
Hart, S., Hart, T.C. and Wright, J.T.
Direct Submission
Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA
Location/Qualifiers
1..8810
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.3-p22.1"
1..8791
/gene="AMELX"
1..1416
/gene="AMELX"
1386..1391
/gene="AMELX"
/join(1417..1472,2781..2846,4783..4830,6240..6281,
6552..6977,8611..8791)
/gene="AMELX"
/product="amelogenin precursor"
1417..1472
/gene="AMELX"
/number=1
2781..2846
/gene="AMELX"
/number=2
/join(2793..2846,4783..4830,6240..6281,6552..6977,
8611..8616)
/gene="AMELX"
/codon_start=1
/product="amelogenin precursor"
/protein_id="AAK77213.1"
/db_xref="GI:15028583"
/translation="MGTWILFACLLGAAPMP,PPHGHGYINFSYEDLYLEAIRID
IRPPSYGVBRMGMLHQIIPVVSQOTPOSHLQPHNHPVPAQDFIQQGPMPLPGHSMPTQHNQNLPLPAQDFPQVQPPQPPHQPQPS
MHPQLPLPMLDPLLEAMPSTDKTKEEVD"
2793..2840
/gene="AMELX"
4783..4830
/gene="AMELX"
/number=3
6107..6148
/gene="AMELX"

gene
promoter
TATA_signal
mRNA
exon
exon
CDS
sig_peptide
exon
exon

	exon	/number=4 6240..6281 /gene="AMELX"
	exon	/number=5 6552..6977 /gene="AMELX"
	exon	/number=6 8611..8791 /gene="AMELX"
	polya_signal	/number=7 8785..8791 /gene="AMELX"
ORIGIN		
Query Match	55.8%; Score 274.6; DB 9;	Length 8810;
Best Local Similarity	79.0%; Pred. No. 2.8e-67;	
Matches 394;	Conservative 0; Mismatches 89; Indels 16; Gaps 5	
Dy	1 ATGGGGAAGCTGCATTTTGTTCCTCCTCGTGAGACAGCTTTGTAAGCCCTGAGT	60
Dy	2793 ATGGGGAAGCTGCATTTTATTTCCTGCCCTCGTGAGAGCACTTTGCAATGCTGAGT	2855
Dy	61 AAAAATACCCCT-----GCATAATATTCATTTTCAAGAAGCTGGAAATTAAGTCTGCC	114
Dy	2853 AAAAACCCCCCTTGCAATGCTGATGCTGCAATTTTCAAAAAGTAAATTAATGCTC	2912
Dy	115 CACAAGTTGTTAAACTTTAGGGTTTAAAGACAGTACAAAGTCAATGCTCTCT	174
Dy	2913 ATAAGTTGGTGAAG--TTAGGGTTTAAACAGATGAGATCAGATGCTCATATGCTCT	2970
Dy	175 GTGTTTAAAGAAACACTGGAAGAGCTGTTTAAAAAATAATATTCAGATGCTCC	234
Dy	2971 GGATTGAAGAAACACTT-CAGAGCTTGTTTAAAAAGTAAATATCTAAATGCCGCTAC	3029
Dy	235 ACCCAAGACTGATTCAGTAGACAGAGTGGGGGAGTGGCCAGACTCTGCATTTTAAC	294
Dy	3030 CAAAATTTCTGATTTTGTATACAGCTGGGGGGG-----GCCAGAGCTCTGCATTTTAT	3083
Dy	295 AAGCACTTCAGAGATTTCTGTGAGACAAATTAATTGTAATATCATGCCCATTCTAG	354
Dy	3084 AAGCACCCCGAGAGATTTCTGTGGAACCTTTAGCTTTGTAATATCACACCACCTCTAG	3143
Dy	355 ATGAGAGAACTTTTAAAGAGGAGCCCTGAAAGGCTCCAGAGAAAGTCTGAAACGCT	414
Dy	3144 ATGAGAGAAAGCTTTTGAAGAGGAGCCCTGAAAGGCTCCAGAGAAAGTCTTAAACGAGCT	3203
Dy	415 TAGG-CAAATACTACAAAAATGCAATTTTCTTAAAAAGCAATTTCTTAAAGAGTGTCCA	473
Dy	3204 TTGGCAAAATATTAACAGAGATGCCAGTTTGTCTTAAACCAATTTCTCTCAAGATTTCCA	3263
Dy	474 ACTCTCTTCCTGAGCCCTCCA 492	
Dy	3264 AATCTCTTCTGCTCTCCA 3282	
RESULT 5		
AL357130/c	158142 bp DNA linear HTG 10-JUL-2001	
LOCUS	Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.	
DEFINITION	AL357130 GI:9863914	
ACCESSION	AL357130.3	
VERSION	HTG; HTGS _PHASE1; HTGS _CANCELLED.	
KEYWORDS	Homo sapiens (human)	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Hominidae; Homo.	
REFERENCE	McClay, K. Direct Submission Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgenet@sanger.ac.uk requests: clonerquest@sanger.ac.uk On Aug. 21, 2000 this sequence version replaced gi:9214076.	
TITLE	JOURNAL	
COMMENT		

```

----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: Bb169E15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 156466 bases at least Q30
Consensus quality: 157038 bases at least Q20
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs quality
coverage: 4.56x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 68897: contig of 68897 bp in length
* 68898 68997: gap of 100 bp
* 68998 122842: contig of 53845 bp in length
* 122843 122943: gap of 100 bp
* 122943 129584: contig of 6642 bp in length
* 129585 129685: gap of 100 bp
* 129685 158142: contig of 28458 bp in length.
Location/Qualifiers
1. 158142
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-169E15"
/clone_1fb="RPC1-13.1"
1. 68897
/note="assembly fragment: 01108
fragment_chain:1
clone_end:17
vector_side:left"
68998. 122842
/note="assembly fragment: 00101
fragment_chain:1"
122943. 129584
/note="assembly fragment: 00361
fragment_chain:1"
129685. 158142
/note="assembly fragment: 01285
fragment_chain:1
clone_end:SP6
vector_side:right"
-----
ORIGIN
Query Match 55.8%; Score 274.6; DB 2; Length 158142;
Best Local Similarity 79.0%; Pred. No. 2.9e-67;
Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTGCTATGCCCCGAGT 60
Db 44325 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTGCTATGCCCCGAGT 44266
QY 61 AAAATACCCCT-----GCATATATTCAATTCACTGGAAGCTGGAATAAAGTCGGCC 114
Db 44265 AAAATACCCCTGGATTAAGTCACTGCTGCAATTTCAACAACCTGGACATTAATAATCTGCTC 44206
QY 115 CACAGTTGTAACTTTAGGGTTTAAAGACAGTACAAGATCAGATGTCTCAATGTCTCT 174

```



```

STS
  /rpt_family="MER20"
  /db_xref="dbSTS:L47508"
  38042..38250
  repeat_region
    /rpt_family="MSTa"
    38122..38510
  repeat_region
    /rpt_family="MSTa"
    38797..40103
  repeat_region
    /rpt_family="MIR"
    complement(42465..42539)
    /rpt_family="MIR"
    47139..47335
  repeat_region
    /rpt_family="MER3"
    complement(48993..49166)
  repeat_region
    /rpt_family="MIR"
    complement(49718..49918)
  repeat_region
    /rpt_family="L1PB1a"
    complement(51943..52505)
  repeat_region
    /rpt_family="L1MB3"
    53226..53677
  repeat_region
    /rpt_family="MLT1d"
    complement(54846..54966)
  repeat_region
    /rpt_family="MIR"
    55040..55226
  repeat_region
    /rpt_family="MER20"
    55691..55837
  repeat_region
    /rpt_family="Alu"
    56051..56486
  repeat_region
    /rpt_family="L1MB3a"
    58557..59030
  repeat_region
    /rpt_family="MLT1d"
    complement(59270..59428)
  repeat_region
    /rpt_family="MER5"
    complement(59847..60030)
  repeat_region
    /rpt_family="MER20"
    complement(61290..61384)
  repeat_region
    /rpt_family="L1PB1"
    61683..61811
  repeat_region
    /rpt_family="MER20"
    61907..62113
  repeat_region
    /rpt_family="L1MB7"
    complement(63588..63969)
  repeat_region
    /rpt_family="L1MA10"
    64800..65073
  repeat_region
    /rpt_family="Alu"
    complement(67157..67209)
  repeat_region
    /rpt_family="MIR"
    67637..75845
  repeat_region
    /gene="HAMELX"
    67637..68496
  repeat_region
    /gene="HAMELX"
    /note="similar to bovine amelogenin promoter"
    join(68495..68552,69861..69926,71863..71910,73320..73361,
    73632..74057,75691..75845)
  repeat_region
    /gene="HAMELX"
    /product="amelogenin"
    join(69873..69926,71863..71910,73320..73361,73632..74057,
    75691..75696)
  repeat_region
    /gene="HAMELX"
    /note="similar to amelogenin encoded by GenBank Accession
    Number M86932"
    Number M86932
    /codon_start=1
  repeat_region
    /product="amelogenin"
    /protein_id="AAC21581.1"
    /db_xref="GI:2739350"
  repeat_region
    /translation="MGTWILFACILGAAPMLPPRHGHGVYINFSYEVLTPLKWTOS
    IRPPYSYGVPMGWLHNOIIPVLSOQHPTTLQPHHIIIPVPAQDPVITQDPMR
    VEGQSMTPRIQHONPLRPAQDQVORQORPQRPMQQRPRVHNMQLPQPPPLP
    MPMQQLPRLPDLITLEAMPSTDKTKREEDV"
  repeat_region
    /rpt_family="MER3"
    complement(77161..77331)
  repeat_region
    /rpt_family="Alu"
    complement(79184..79473)
  repeat_region
    /rpt_family="Alu"
    complement(79881..80169)
  repeat_region
    /rpt_family="Alu"

```

```

  repeat_region 80196..80253
    /rpt_family="MER20"
  repeat_region 80384..80459
    /rpt_family="L1PB3"
  repeat_region complement(81544..81633)
    /rpt_family="MIR"
  repeat_region 84789..84957
    /rpt_family="L1MA2"
  repeat_region 84978..85252
    /rpt_family="Alu"
  repeat_region 85294..85692

Query Match      55.8%; Score 274.6; DB 9; Length 259202;
Best Local Similarity 79.0%; Pred. No. 2.9e-67;
Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;

QY 1 ATGGGACCTGATTTTGTGGCTGCTCCGAGGAGCAGCTTTGCTATGCCCCGAGT 60
Db 69873 ATGGGACCTGATTTTGTGGCTGCTCCGAGGAGCAGCTTTGCTATGCCCCGAGT 69932
QY 61 AAAATACCCCT-----GATAATATTCATTTACAGCTTGAATTAAGTCTGCC 114
Db 69933 AAAACACCCCTGACATTAAGTCTGATTCATTAATTCATTAATTAAGTCTGCC 69992
QY 115 CACAGTTGGTAACTTTAGGTTTAAAGACATTAAGATCAGATGCTCAATGCTCT 174
Db 69993 ATAGTTGGTAA--TTAGGTTTAAACATTAAGATCAGATGCTCTCAATGCTCT 70050
QY 175 GTGTTTAAAGAACACTTGGAGAGCTTTGATTAATAATAATATTCAGATGCTCC 234
Db 70051 GGTGTAAGAAACACTT--CAGAGCTTTGTTTAAAGATTAATTCAGATGCTCC 70109
QY 235 ACCCAAGATGATTCAGTACGAGACAGAGTGGGGAGAGTCCAGACCTGATTTAAC 294
Db 70110 CAAAATTCGATTTGATTCAGTACGAGTGGGGAGAGTCCAGACCTGATTTAAC 70163
QY 295 AAGCACTCAGAGATTCGTGAGACATTAATTAATATTCAGTCCCATCTCTAG 354
Db 70164 AAGCACTCAGAGATTCGTGAGACATTAATTAATATTCAGTCCCATCTCTAG 70223
QY 355 ATGAGAGAACTTTTGAAGGAGACCTTGAAGGCTCCAGAGAAAGTCTGAAACAGT 414
Db 70224 ATGAGAGAACTTTTGAAGGAGACCTTGAAGGCTCCAGAGAAAGTCTGAAACAGT 70283
QY 415 TTGG--CAATATCTACAAAATGCAATTTCTCTAAACCAATTTCTAACAGTCTCA 473
Db 70284 TTGGACAAATATCTACAGATGCAATTTCTCTAAACCAATTTCTAACAGTCTCA 70343
QY 474 ACTCTCTTCTGCTCCCTCCA 492
Db 70344 AATCTCTTCTGCTCCCTCCA 70362

RESULT 7
AB091781 6465 bp DNA linear Pri 02-MAY-2003
LOCUS Pan troglodytes AMELX gene for amelogenin, partial cds.
DEFINITION AB091781
ACCESSION AB091781
VERSION AB091781.1 GI:29126015
KEYWORDS
SOURCE
  ORGANISM
    Pan troglodytes (chimpanzee)
    Pan troglodytes
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
REFERENCE
  1 Iwase,M., Satra,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
    From the Cover: The amelogenin loci span an ancient pseudautosomal
    boundary in diverse mammalian species
    Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
JOURNAL MEDLINE
PUBMED 12672962
REFERENCE 2 (bases 1 to 6465)
AUTHORS Iwase,M., Satra,Y. and Takahata,N.

```

TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES

source

Location/Qualifiers
1. .6465
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/sex="male"
927. .6465
/gene="AMELX"
/join(927. .982,2283. .2294)
/gene="AMELX"
/join(2295. .2348,4284. .4331,5607. .5648,5740. .5781,
6052. .>6465)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66101.1"
/db_xref="GI:29126016"
/translation="MGTWILFACILGAAPAMP.LPHGHGPGYINFSYENSHQAINVD
RTALVLPKWKOSIRPVPYSGYBPGMGLHNOIIPVLSQOHPRHTLQPHHPIPV
PAQOPVIRQOPMVRPQGSHMTPIQHNPRLPRPAQOPVQOPVQOPMOPORPV
HMPORLPORPLPMPMPORPLPMLPRLTLEAMPSIDKX"

ORIGIN

Query Match 55.5%; Score 273; DB 9; Length 6465;
Best Local Similarity 78.8%; Pred. No. 7,9e-67;

Matches 393; Conservative 0; Mismatches 90; Indels 16; Gaps 5;

QY 1 ATGGGAGCTGGATTTTGTTCCTGCTGCTGAGACACCTTGTCTATGCCGTGAGT 60
DB 2295 ATGGGAGCTGGATTTTGTTCCTGCTGCTGAGACACCTTGTCTATGCCGTGAGT 2354
QY 61 AAAATACCC-----TGCAATATTCATTTTCAACAAGCTTGAATAAAGTCGCC 114
DB 2355 AAAACACCCCTGCATATGTCAGTCCAAATTCACAAAGCTTGAATAAAGTCGCC 2414
QY 115 CACAGTTGTAACCTTAGGGTTTAAAGACAGTACAGATCAGATGTCCTCAATATGTCCT 174
DB 2415 ATAGTTGGTAAA--TTAGGTTTAAACAGTATGATCAGATGTCCTCAATATGTCCT 2472
QY 175 GTGTTAAGAAACACTTGAAGAGCTTGTATTAATAAATAATATATCCAGATGCTCC 234
DB 2473 GGGTTGAAGAAACCTT--CAGAGACTTGTTTAAAGATATATTCACAAATGCCGTAC 2531
QY 235 ACCCAAGACTGATTCAGTACAGACAGAGTGGGGAGAGTCCCAAGACTTGCATTTTAA 294
DB 2532 CAAAATTCGATTTGGTACAGTGGGGCGGG-----GCCCAAGACTTGCATTTTAT 2585
QY 295 AAGCACTCAGAGATCTGTGAGACATTAATCTTAATATCATGCGCCATCTCTAG 354
DB 2586 AAGCACTCAGAGATCTGTGAGACATTAATCTTAATATCATGCGCCATCTCTAG 2645
QY 355 ATGAGAGAACTTTTGAAGAGACCTTGAAGAGCTTCAAGAAAGTGTCTGAACAGCT 414
DB 2646 ATGAGAGAACTTTTGAAGAGACCTTGAAGAGCTTCAAGAAAGTGTCTGAACAGCT 2705
QY 415 TAGG-CAAATACATCAAAAATGCAATTTCTCTAAACCCCAATTTTAAAGAGTCCCA 473
DB 2706 TTGGACAAATATTAACAGATGCGAGTTTGTCTAAACCCCAATTTCTCTCAAGATTCCA 2765
QY 474 ACTCTCTCTGCCCTCCA 492
DB 2766 AATCTCTCTGCCCTCCA 2784

RESULT 8
AB091785
LOCUS AB091785 5684 bp DNA linear PRI 02-MAY-2003

DEFINITION Lemur catla AMELX gene for amelogenin, partial cds.
ACCESSION AB091785
VERSION AB091785.1 GI:29126023
KEYWORDS
SOURCE
ORGANISM
Lemur catla (ring-tailed lemur)
Lemur catla
Mammalia; Eutheria; Primates; Strepsithini; Lemuridae; Lemur.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@minokoryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES

source

Location/Qualifiers
1. .5684
/organism="Lemur catla"
/mol_type="genomic DNA"
/db_xref="taxon:9447"
100. .5684
/sex="male"
/gene="AMELX"
/join(100. .156,1503. .1514)
/gene="AMELX"
/join(1515. .1568,3465. .3512,4804. .4845,4937. .4981,
5229. .>5684)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66105.1"
/db_xref="GI:29126024"
/translation="MGTWILFACILGAAPAMP.LPHGHGPGYINFSYENSHQAINVD
RTALVLPKWKOSIRPVPYSGYBPGMGLHNOIIPVLSQOHPRHTLQPHHPIPV
PAQOPVIRQOPMVRPQGSHMTPIQHNPRLPRPAQOPVQOPVQOPMOPORPV
HMPORLPORPLPMPMPORPLPMLPRLTLEAMPSIDKX"

ORIGIN

Query Match 53.0%; Score 261; DB 9; Length 5684;
Best Local Similarity 78.4%; Pred. No. 2.1e-63;

Matches 393; Conservative 0; Mismatches 90; Indels 18; Gaps 6;

QY 1 ATGGGAGCTGGATTTTGTTCCTGCTGCTGAGACACCTTGTCTATGCCGTGAGT 60
DB 1515 ATGGGAGCTGGATTTTGTTCCTGCTGCTGAGACACCTTGTCTATGCCGTGAGT 1574
QY 61 AAAATACCCCT-----GCATATATTCATTTTCAACAAGCTTGAATAAAGTCGCC 114
DB 1575 AAAACACCCCTGCATATGTCAGTCCAAATTCACAAAGCTTGAATAAAGTCGCC 1634
QY 115 CACAGTTGTAACCTTAGGGTTTAAAGACAGTACAGATCAGATGTCCTCAATATGTCCT 174
DB 1635 CACCGCTGTAACCTTAGGGTTTAAAGATGATATATCATGATGTCCTCAACGTCCTC 1694
QY 175 GTGTTAAGAAACACTTGAAGAGCTTGTATTAATAAATAATATATCCAGATGCTCC 234
DB 1695 GTGTTAAGAAACACTT--GAGAGATCGTTTAAAGATTAAGATTCCTCAATGCTCTGC 1753
QY 235 ACCCAAGACTGATTCAGTACAGACAGAGTGGGGAGAGTCCCAAGACTTGCATTTTAA 294
DB 1754 CAAAATTCGATTTGGTACAGCTGAA-----GCCAGGCGCAAGACTTGCATTTTAA 1807
QY 295 AAGCACTCAGAGATCTGTGAGACATTAATCTTAATATCATGCGCCATCTCTAG 354
DB 1808 AAGCACTCAGAGATCTGTGAGACATTAATCTTAATATCATGCGCCATCTCTAG 1867

	AB091783	6442 bp	DNA	linear	PRI 02-MAY-2003
LOCUS	AB091783				
DEFINITION	Saimiri sciureus AMELX gene for amelogenin, partial cds.				
ACCESSION	AB091783				
VERSION	AB091783.1	GI:29126019			
KEYWORDS					
SOURCE	Saimiri sciureus (common squirrel monkey)				
ORGANISM	Saimiri sciureus				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
22608569
12672962
2 (bases 1 to 6442)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (13-SEP-2002) Mineo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science,
Shonan kohnasimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@minskoryu01.soken.ac.jp, Tel:81-468-58-1571,
Fax:81-468-58-1544)

	FEATURES	source	Location/Qualifiers
			1..6442
			/organism="Salimiri sciureus"
			/mol_type="genomic DNA"
			/db_xref="taxon:9521"
			/sex="male"
	gene		936..6442
			/gene="AMELX"
	5' UTR		join(936..991,2298)..2309)
			/gene="AMELX"
CDS			join(2310..2263,4289..4336,5592..5633,5725..5769, 6041..56442)
			/gene="AMELX"
			/codon_start=1
			/product="amelogenin"
			/protein_id="BAC66103.1"
			/db_xref="GI:29126020"
			/translation="MGTWILPACLLGAAPMLPRPHGPGHYINESYENSHQAINID RTGVLTPLKYSQSMIRAIRPSYSYERMGSLNHQITPVLSQQHPPTHTLQSFNHIV VPAAQVVFQDFPMFVPEQASHMITQNHPRLPPAQQPQPQVTFVQFPQFPQPP VHPHQPLRPQRLLPRMFMORLPMLRDLPLEAMPAI"

Query Match	52.2%	Score 256.6	DB 9	Length 6442
Best Local Similarity	79.3%	Pred. No. 3,86-62		
Matches	394	Conservative	0	Mismatches 89
			Indels	14
			Gaps	7
QY	1	ATGGGACACCGAATTTGTTGCTCCGCTCCTCGGAGACAGCCTTTCCTATAGCCCTGAGT	60	
	2310	ATGGGACCTGATGATTTTGTTCCTCGCTCCTCGGAGACAGCTTTTGCAATGCTCTGTAGT	2369	
QY	61	AAAA-----TACCCCTGCATTAATTAATTCATATTTACACAGCTTGGAAATAAAGCTGCCCCA	116	

Db	2370	AAACCCTTCATATGACATGCTCAATTCATATAGTTAGCAATATAAATCTGCCCCA	2425
Qy	117	CAGTTGGTAAACCTTTAGGGTTTAAAGACATGACAAAGATCATGTCCTCAAAATGTCCTGT	176
Db	2430	TGGTTGGGAA-GTTAGGGTTTAAACAGATGAGATAGATAGATGCTTCMAACATTTCTGG	2488
Qy	177	GTTTAAAGAAACACTTGGAAAGACTTGTATATATAAAAAATATATTCAGATGGCTCCAC	236
Db	2489	GTTAGAGAAACATT-CAGAGCTTGTTTT--AAAAAGATGATTTCTCAAAATGCTCCAG	2545
Qy	237	CCAAAGTGAATTCAGTAAAGCAGAGATGGGGGAGTGGCCAGACTCTGCATTTTAACAA	296
Db	2546	CAAA-----GATTCGATTTGGTACAGCTGGGGGTGGGGCCAGAGACTGCATTTTATA	2601
Qy	297	GCACCTCAGAGATTTCTGTGAGACATTTAACTTGTAAATTCATCCGCCACTCTCAT	356
Db	2602	GTTACACAGAGATTCGTGTAAACCAATAGCTGTATGATTCACACACCACATCTCTGAT	2661
Qy	357	GGAGGAAACTTTTGAAGAGAGACCCCTGTAAGAGCCTCAGAGAAAGTGTCCAAAGC-TT	415
Db	2662	GGAGGAGACTTTTGAAGAGAGACCTTGAAAGTCTTAGAGAAAGTGTTCACACAGCTT	2721
Qy	416	AGGCAATACTACAAAAATGCAATTTTCTTAATAACCAATTTCTTAACGAGTGTCCAC	475
Db	2722	GGGCAAAATATTACAGAGATGCAAGTTTGTCTAAAA--CAATTCCTCTCAAGTTCCAAA	2780
Qy	476	TCTCTTCCTGCTCCCA	492
Db	2781	TATCTTCTGCTCTCCA	2797

RESULT 10				
LOCUS	AB091794	5591 bp	DNA	linear
DEFINITION	Equus caballus AMELY gene for amelogenin, partial cds.			
ACCESSION	AB091794			
VERSION	AB091794.1	GI:29126040		
KEYWORDS				
SOURCE				
ORGANISM	Equus caballus (horse)			
	Equus caballus			
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.			
REFERENCE	1			
AUTHORS	Iwase,M., Satta,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.			
TITLE	From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)			
MEDLINE	22608569			
PUBMED	12672962			
REFERENCE	2 (bases 1 to 5591)			
AUTHORS	Iwase,M., Satta,Y. and Takahata,N.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies(Sokenkai), Department of Biogstems Science; Shonan kokuuimura, Hayama, Kanagawa 240-0193, Japan (E-mail:iwasemin@koryu.w01.soken.ac.jp, Tel:81-468-1571, Fax:81-468-58-1544)			

FEATURES	Location/Qualifiers
source	1..5591
	/organism="Equus caballus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9796"
	/sex="male"
gene	130..5591
	/gene="AMELY"
5'UTR	join(130..180,1483..1494)
	/gene="AMELY"
CDS	join(1495..1548,3470..3517,4907..4915,5235..5559)
	/gene="AMELY"
	/codon_start=1
	/product="amelogenin"
	/protein_id="BAC66113.1"

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Sekiuchi, H., Minaguchi, K., Machida, Y. and Yakushiji, M.
PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta
Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)
99234629
10218009
2 (bases 1 to 1935)
Sekiuchi, H.
Direct Submission
Submitted (29-FEB-1996) Hiroshi Sekiuchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Mihama-ku, Chiba, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)
Location/Qualifiers
1. 1935
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
1. 1935
/gene="AMGX"
1. 1935
/gene="AMGX"
/note="amelogenin
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"
/number=2

ORIGIN

Query Match 44.6%; Score 219.6; DB 9; Length 1935;
Best Local Similarity 78.5%; Pred. No. 1.4e-51;
Matches 328; Conservative 0; Mismatches 79; Indels 11; Gaps 5;

QY 76 AATATCAATTTTCAAGCTTGAATTAAGTCTGCCCAAGTGTGTAACCTTTAGG 135
Db 28 AGTCCCAATTTTCAAGCTTGAATTAAGTCTGCCCAAGTGTGTAACCTTTAGG 85

QY 136 TTAAAGACATGACAGATGATGCTCAATGCTCTGTGTTTAAAGAACTTTAGG 195
Db 86 TTAAAGACATGACAGATGATGCTCAATGCTCTGTGTTTAAAGAACTTTAGG 144

QY 196 GAGCTTTTAAAGAACTTTTAAAGTCTGCCCAAGTGTGTAACCTTTAGG 255
Db 145 GAGCTTTTAAAGAACTTTTAAAGTCTGCCCAAGTGTGTAACCTTTAGG 204

QY 256 GCAAGAGTGGGGAGTGGCCAGAGCTCTGCAATTTTAAAGCACTCAGAGATTCGT 315
Db 205 GCTGGGGGGG-----GCCAGAGCTCTGCAATTTTAAAGCACTCAGAGATTCGT 258

QY 316 GGAGACATTTAATCTTGAATATATCATGCCCCATCTTGAATGAGAACTTTAGAG 375
Db 259 TGAAGATCTTGAATATATCATGCCCCATCTTGAATGAGAACTTTAGAG 318

QY 376 GACCTTGAAGAGCTCTGAGAAAGTCTGCAAGCTTGAAG--CAATCTCAAAAT 434
Db 319 GACCTTGAAGAGCTCTGAGAAAGTCTGCAAGCTTGAAG--CAATCTCAAAAT 377

QY 435 GCCAATTTTCTTAAAGCCCAATTTTAAAGAGTCTGCAAGCTTCTGCTCCCA 492
Db 378 GCCAATTTTCTTAAAGCCCAATTTTCTGCAAGATTCCTTCTGCTCCCA 435

RESULT 13
AB091790 6264 bp DNA linear MAM 02-MAY-2003
LOCUS Bos taurus AMELY gene for amelogenin, partial cds.
DEFINITION
ACCESSION AB091790
VERSION AB091790.1 GI:29126032
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1
Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
22608569
12672962
2 (bases 1 to 6264)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submission
Submitted (19-SEP-2002) Mineo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science, Shonan Kokusai-mura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase@nkcoryu.w01.soken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)
Location/Qualifiers
1. 6264
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/sex="male"
114. 6264
/gene="AMELY"
/gene="AMELY"
/join(114. 169, 1585. 1596)
/gene="AMELY"
/join(1597. 1659, 3515. 3562, 5080. 5124, 5863. 56264)
/gene="AMELY"
/codon_start=1
/product="amelogenin"
/protein_id="BA066109.1"
/db_xref="GI:29126032"
/translation="MGWTLFACLLGAGVMPVSIIPHPHPGVINSEYVLPKX
YONMLRYPSPSGYBEPVGMHROIIPVVSQSPQNHAIQPHHNPVPAQDPVPOQ
PMVPGVGHSMPIQIHQCPNLPPLAQGSFQPPPIOPPPPPVHPHPIORLPPOP
PPIPPMPQPLPVLPLPLAMPAT"

ORIGIN

Query Match 33.8%; Score 166.2; DB 4; Length 6264;
Best Local Similarity 68.0%; Pred. No. 2.4e-36;
Matches 338; Conservative 0; Mismatches 118; Indels 41; Gaps 6;

QY 1 AATGAGACCTGGAATTTGTTGCTGCTCCGCGGAGCACTTGTGTAAGCCGTAGT 60
Db 1597 AATGAGACCTGGAATTTGTTGCTGCTCCGCGGAGCACTTGTGTAAGCCGTAGT 1656

QY 61 AATATACCCCTGCATATATTCATTTTCAAGCTTGAATTAAGTCTGCCCAAGT 120
Db 1657 AATATACCCCTGCATATATTCATTTTCAAGCTTGAATTAAGTCTGCCCAAGT 1708

QY 121 TGGTAACTTTAGGTTTAAAGACATGATCAATGCTCTCAATGCTCTGTTT 180
Db 1709 AGGTAACTTTAGGTTTAAAGACATGATCAATGCTCTCAATGCTCTGTTT 1763

QY 181 AAGAAACCTTGAAGAGCTTGTAT--AAAAAATATATTTCCAGATGCTCCAC 237
Db 1764 AAGATCAATTTGAAGAGCTTGTATTTAAAAAAGATCCCAAGATTTTCTGCT 1823

QY 238 CAAGA--CTGATTCAGTGAAGAGAGTGGGGAGTGGCCAGAGCTTGCAATTTTAA 295
Db 1824 AAGATTCAGTGAAGAGAGTGGGGAGTGGCCAGAGCTTGCAATTTTAA 1861

QY 296 AGCACTCAGAGATTTCTGAGACATTTAATTTGTAATATCATGCCCCATCTGAGA 355
Db 1862 AAGATTCAGAGATTTCTGAGACATTTAATTTGTAATATTTAATTTTAAAT 1921

QY 356 TGAAGAACTTTTGAAGGAGCCCTTGAAGGCTCCAGAGAAAGTCTGCAAGCTT 415
Db 1922 TGAAGAACTTTTGAAGGAGCCCTTGAAGGCTCCAGAGAAAGTCTGCAAGCTT 1980

QY 416 AGCAAAATCTACAAAATGCCAATTTTCTTAAACCAATTTTCAAGAGTGTCCAC 475

```

Db      1981 TAGTAAATACAAATATCTAGTTTACTATAAATCCAAITTTCTTACAGATTATTA 2040
QY      476 TCTCTCTGCTGCTCCA 492
Db      2041 TATTTCTGCTGCTCCA 2057

RESULT 14
AF294397      9384 bp  DNA  linear  ROD 03-SEP-2000
DEFINITION   Mus musculus amelogenin gene, promoter and partial cds.
ACCESSION    AF294397
VERSION      AF294397.1 GI:9965405
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    Sneed,M.L., Paine,M.L., Luo,W., Zhu,D.H., Yoshida,B., Lei,Y.P.,
AUTHORS      Paine,C.T., Chen,L.S., Burslein,J.M., Jilpukdeebudintira,S.,
              White,S.N. and Birlings,P. Jr.
              Transgene animal model for protein expression and accumulation into
              forming enamel
              Connect. Tissue Res. 38 (1-4), 279-286 (1998)
JOURNAL      20515040
MEDLINE      11063035
REFERENCE    2 (bases 1 to 9384)
AUTHORS      Sneed,M.L., Zhu,D.-H., Lei,Y.-P. and Paine,M.L.
TITLE        Direct Submission
JOURNAL      Submitted (08-AUG-2000) Dentistry, University of Southern
              California, 2250 Alcazar Street, CSA103, Los Angeles, CA 90033, USA
FEATURES
  source
    1. .9384
       /organism="Mus musculus"
       /mol_type="genomic DNA"
       /db_xref="taxon:10090"
       /chromosome="X"
       1..8546
       promoter
       TATA_signal
       mRNA
       CDS
         8547..>8612
         /product="amelogenin"
         8559..>8612
         /note="enamel matrix protein"
         /codon_start=1
         /product="amelogenin"
         /protein_id="AAG10078.1"
         /db_xref="GI:9965406"
         /translation="MGTWILFACLLGAAFPMP"

ORIGIN
Query Match      32.6%; Score 160.6; DB 10; Length 9384;
Best Local Similarity 65.7%; Pred. No. 9,7e-35;
Matches 332; Conservative 0; Mismatches 154; Indels 19; Gaps 6;

QY      1 ATGGGACCTGGATTTTGGCTGCTCTGCGAGACACCTTTGCTATGCCCGTAGT 60
Db      8559 ATGGGACCTGGATTTTGGCTGCTCTGCGAGACACCTTTGCTATGCCCGTAGT 8618
QY      61 AAATATACCC-----TGCATATATTTCAATTTTCAGAGCTTGGAAATTAAGTGTG-CC 114
Db      8619 AAATATATCATTTTACTTACATTTGCAATTCATTAATCTTGGAACTAGAAATCTGCTCC 8678
QY      115 CACAGTTGTAACCTTTAGGCTTTTAAAGACGTAACAAGATCAGATGCTCCTCAATGCTCT 174
Db      8679 CAAAGTGTGTAATTTTCACTGTGTAAACAGTCAAGATCTTATATCCCAAGATCTAA 8738
QY      175 GTGTTTAAAGAAACCTTGAAGAGCTGTGTATTAATAAATAATATTTCCAGATGCTCC 234
Db      8739 TAGGTGTGAAGAAATATTTTGAAGTTTGTAAAAAGAA-----ATTCAGAGAAAGCTCC 8794
QY      235 ACCCAAGAC--TGATTCAGTAGAGACAGAGTGGGGGAGTGGCCAGGACTCTGCATTTA 292

```

```

Db      8795 ACCTTAGACTGTATTCAGTTTCAGTGAAGTGGACCAAGCACTCAAGATTTTAACAGC 8854
QY      293 ACAAGACCTCAGAGATTTCTGTGAGACAAATTATCTGTAAATATCATGCCATCTCT 352
Db      8855 ATTTGACCCCTAGAGATCTTCTGTAAAGATAGTTGTAACAAACATCGCCATCTCT 8914
QY      353 AGATGAGGAAATTTTAAAGGAGACCTTGAAGAGCTTCAGAGAAAGTGTCTGAACAG 412
Db      8915 ACAAGAGATATATCTTATTTGAAGGAGACCTTGAAGAGCTTCAGAG--AGTGTACTGT 8972
QY      413 CTAGGCAATATACAAAATGCG-----ATTTCTGTAAACCAATTTCTAAGAG 467
Db      8973 TTAGGCAATATTAACAGCCCCCCCCCAAAAATCTTAAGCTTAATTACTCAGAG 9032
QY      468 TGTCAACTCTTCTGCTGCTCCA 492
Db      9033 TTCCAAATTTCTTCTGCTGCTCCA 9057

RESULT 15
AL805974      95826 bp  DNA  linear  ROD 29-AUG-2003
LOCUS        AL805974/c
DEFINITION   Mouse DNA sequence from clone Rp23-334F21 on chromosome X, complete
              sequence.
ACCESSION    AL805974
VERSION      AL805974.8 GI:34366495
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 95826)
AUTHORS      Chapman,J.
TITLE        Direct Submission
JOURNAL      Submitted (28-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquerry@sanger.ac.uk
              On Aug 30, 2003 this sequence version replaced gi:25955748.
              Sequence from the Mouse Genome Sequencing Consortium whole genome
              shotgun may have been used to confirm this sequence. Sequence data
              from the whole genome shotgun alone has only been used where it has
              a phred quality of at least 30.
              ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              -----

COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMBPP; Information
on the WORMBPP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormbep Rp23-334F21 is
from the RPI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: PBAC3.6.
Location/Qualifiers

```

Source

```
1. . 95826
/organism="Mus musculus"
/moi_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-334F21"
/clone_lib="RPCI-23"
```

ORIGIN

Query Match	32.6%;	Score 160.6;	DB 10;	Length 95826;
Best Local Similarity	65.7%;	Pred. No. 1e-34;		
Matches 332;	Conservative	0;	Mismatches 154;	Indels 19;
				Gaps 65;

QY	1	ATGGGGACCTGGAAATTTTGTGTGGCTGCTCTGCGGAGCAGCCCTTGCTATGCCCCGGAGT	60
Db	22703	ATGGGGACCTGGAAATTTTGTGTGGCTGCTCTGCGGAGCAGCCCTTGTCTATGCCCCGGAGT	22644
QY	61	AAAATATCCC----TGCATATATTTCAATTTTCAACAAGCTTGGAAATTAAGTTCG-CC	114
Db	22643	AAAACATCATTTTACTTAATTTGATCCAAATTAATATATCTTGGAACTGAAGATCTGCTCC	22584
QY	115	CACAGTGTGAATCTTAGGGTTTAAAGACAGTACCAAGATCAGATGCTCTCAATGCTCT	174
Db	22583	CATAGTTGGTATTTTCACTGTGTAAACAGTGCAGATCTTATATCCCCAAGTATCTAA	22524
QY	175	GTTGTTAAGAAACCTTGGAAAGCTTGTATATAAAAAAATATATTTCCAGATGCTCC	234
Db	22523	TAGGTGTTGAAGAAATCATTTTGAAGTTGTATAAAAAG--AATCAGGAAAGTCTCC	22468
QY	235	ACCCAAGAC--TGATTCAGTGAAGACAGAGTGGGGGAGTGCCCAAGACTCTGCATTTTA	292
Db	22467	ACCTTAACCTGTGATTCAGTTCACTGAAATGGGACCAACACTCAAGATTTTAAACAAG	22408
QY	293	ACAAGCACCTCAGAGATTCGTGGAGACAATTACTGTAAATATCATGCCCCACTCT	352
Db	22407	ATTTGACCCCTAGAGAGATCTTCTCTGTACATTAAGCTGTGAAACAACACTGCCCACTCT	22348
QY	353	AGATGGAGGAACCTTTAGAAGGGACCTTGAAAGGCTCCAGAGAAAGTCTCGAACAG	412
Db	22347	ACAAGTATGTTATCTAATTTGAAGGGACCTTGAAAGGCTTCCAGAG--AGTGTTACTGT	22290
QY	413	CTTAGGCAAAATACATAAAAAATGCCC---AATTTTCTTAAACCACATTTCTTAACGAG	467
Db	22289	TTTAGGCAAAATATATAACAGCCCCCCCCCAAAAAAATCTTAAGGCTTAATTACTCAGAAG	22233
QY	468	TGTCCAACTCTCTTCGCCCCCTCA	492
Db	22229	TTCCAAATATTTCTTCTGGCCCTCA	22205

Search completed: February 8, 2005, 16:37:05
Job time : 2745.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 652.659 Seconds
(without alignments)
3957.220 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492
Sequence: 1 atgsgagaccgagcttctgtc.....aactctctctgcctcca 492

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001a:*
5: geneseqn2001b:*
6: geneseqn2002a:*
7: geneseqn2002b:*
8: geneseqn2003a:*
9: geneseqn2003b:*
10: geneseqn2003c:*
11: geneseqn2003d:*
12: geneseqn2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274.6	55.8	259202	12	ADQ18492
2	51.4	10.4	270	12	ADM80835
3	51.4	10.4	549	12	ADM80836
4	51.4	10.4	623	12	ADM80827
5	51.2	10.4	476	3	AAZ50832
6	50.2	10.2	556	10	ADBS9026
7	50.2	10.2	556	10	ADBS3782
8	49.8	10.1	722	6	AA14111
9	49.8	10.1	752	6	AA14110
10	48.2	9.8	318	12	ADM80836
11	48.2	9.8	750	2	AAZ07020
12	48.2	9.8	750	2	AAZ07020
13	46.6	9.5	802	12	AAZ07019
14	46.6	9.5	852	12	ADQ22977
15	45.2	9.2	65	6	ABN28389
16	39.8	8.1	95000	12	ADOS6276
17	39.4	8.0	216215	10	ADP69167
18	39.2	8.0	97658	8	ABQ83210_3
19	37.4	7.6	5310	4	AAK65863
20	36.2	7.4	595	12	ACH74683
21	36.2	7.4	166910	12	ADN01278

C	22	35.8	7.3	374	2	AAV78614	AAV78614	Staphyloc
C	23	35.8	7.3	742	2	AAV7973	AAV7973	Staphyloc
C	24	35.8	7.3	1468	8	ADA89829	ADA89829	Staphyloc
C	25	35.8	7.3	1665	8	ACF74621	ACF74621	Staphyloc
C	26	35.8	7.3	10758	2	AAV74532	AAV74532	Staphyloc
C	27	35.6	7.2	7536	10	ADBS3812	ADBS3812	Human pro
C	28	35.2	7.2	25001	10	ADC87256	ADC87256	Human GPC
C	29	34.8	7.1	2282	4	AAH17568	AAH17568	Human CDN
C	30	34.8	7.1	14252	4	AA104590	AA104590	Human rep
C	31	34.8	7.1	14252	4	AA104692	AA104692	Human rep
C	32	34.8	7.1	14252	4	AB197599	AB197599	Human tes
C	33	34.8	7.1	14252	4	AB197513	AB197513	Human tes
C	34	34.8	7.1	16545	6	AB132050	AB132050	Human tes
C	35	34.8	7.1	24200	4	AB121182	AB121182	Human tes
C	36	34.8	7.1	24206	4	AB121180	AB121180	Drosophila
C	37	34.8	7.1	24235	4	AB105026	AB105026	Drosophila
C	38	34.8	7.1	300000	10	ADBS6352	ADBS6352	Human PTP
C	39	34.8	7.1	300001	12	ADQ14076	ADQ14076	Human pro
C	40	34.6	7.0	837	4	AAH73049	AAH73049	Human cer
C	41	34.6	7.0	1878	10	ACC61054	ACC61054	Gene sequ
C	42	34.6	7.0	1878	10	ADK62759	ADK62759	Disease t
C	43	34.6	7.0	2000	8	ADA68928	ADA68928	Atrialdops
C	44	34.2	7.0	1668	8	ACA46163	ACA46163	Prokaryot
C	45	34.2	7.0	1668	8	ACA46163	ACA46163	Prokaryot

ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
ID	ADQ18492;	
XX	ADQ18492;	
XX	26-AUG-2004 (first entry)	
DT	26-AUG-2004 (first entry)	
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
DE	Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KW	Homo sapiens.	
XX	10-JUN-2004.	
XX	26-NOV-2003; 2003MO-US038193.	
PF	26-NOV-2002; 2002JUS-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PA	Aziz N, Glinesburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
DR	Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.	
PT	Example 2; SEQ ID NO 1311; 210bp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual, and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue	

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 55.8%; Score 274.6; DB 12; Length 259202;

Best Local Similarity 79.0%; Pred. No. 1.6e-73;

Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;

QY 1 ATGGGACCTGATTTTGTGCTGCTGCGAGAGAGGCTTGTGATGCGCTGAGT 60

Db 69873 ATGGGACCTGATTTTGTGCTGCTGCGAGAGAGGCTTGTGATGCGCTGAGT 69932

QY 61 AAAATACCCCT-----GCATTAATTCATTTTCAAGCTTGGAAAATAAGTCTGCC 114

Db 69933 AAAACACCCCTTGCATTAAGTGTCCATTTTCAAACTTGGACATTAATAATCTGCTC 69992

QY 115 CACAGTTGTAAGCTTTAGGGTTAAGCAGTACAAAGTCAATGCTCTCAATGCTCT 174

Db 69993 ATAGTTGTTGTAAGCTTTAGGGTTAAGCAGTACAAAGTCAATGCTCTCAATGCTCT 70050

QY 175 GTGTTTAAAGAACTTGGAGAGCTTGTATTAATAAATAATATATCCAGATGCTCC 234

Db 70051 GGGTTGAAGAACTT--CAGAGCTTGTATTAATAAATAATATATCCAGATGCTCC 70109

QY 235 ACCCAAGCTGATTCAGTAAAGCAGAGTGGGGAGAGTGGCCAGACTCTGCATTTTAC 294

Db 70110 CAAAATTTGATTTTGTAAAGCAGTGGGGAGAGTGGCCAGACTCTGCATTTTAT 70163

QY 295 AAGCACTCAGAGATCTGTGAGAGCAATTAATTGTAATATCATGCGCCATCTCTAG 354

Db 70164 AAGCACTCAGAGATCTGTGAGAGCAATTAATTGTAATATCATGCGCCATCTCTAG 70223

QY 355 ATGAGAGAACTTTTAAAGAGGACCTTGAAGGCTTCCAGAGAAAGTGTCTGACAGCT 414

Db 70224 ATGAGAGAACTTTTAAAGAGGACCTTGAAGGCTTCCAGAGAAAGTGTCTGACAGCT 70283

QY 415 TAGG-CAAAATCTCAAAAATGCAATTTTCTTAAACCAGATTTTAAAGAGTCTCA 473

Db 70284 TTGAGCAAAATTTTCAAGAGATGCGAGTTTGTCTTAAACCAGATTTTCAAGAGTCTCA 70343

QY 474 ACTCTCTCTGCGCTCTCA 492

Db 70344 AATCTCTCTGCGCTCTCA 70362

RESULT 2

ID ADM80835 standard; cDNA; 270 BP.

XX ADM80835;

DT 03-JUN-2004 (first entry)

DE Human CADECM-22 encoding cDNA SEQ ID NO:64.

XX human; cell adhesion and extracellular matrix protein; CADECM;
 KM neuroprotective; cytoskeletal; anorectic; immune disorder;
 KM neurological disorder; developmental disorder;
 KM connective tissue disorder; cell proliferative disorder; cancer; obesity;
 KM tangler disease; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 36..212

FT /*tag= a

FT /product= "CADECM-22"

PN MO2004015396-AA2.

XX 19-FEB-2004.

PF 12-AUG-2003; 2003MO-US025418.

XX 13-AUG-2002; 2002US-0403781P.

PR 30-AUG-2002; 2002US-0407034P.

PR 13-SEP-2002; 2002US-0410566P.

PR 24-SEP-2002; 2002US-0413482P.

PR 25-SEP-2002; 2002US-0413890P.

PR 08-NOV-2002; 2002US-0424904P.

PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;

PI Becha SD, Margulis JP, Swarnaker A, Chawla NK, Raskumar J;

PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;

PI Wang JT, Chien D, Yang YG;

XX WPI; 2004-191795/18.

DR P-PADB; ADM80793.

XX New cell adhesion and extracellular matrix proteins, useful in

PT diagnosing, treating and preventing immune, neurological, developmental,

PT connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 64; 272bp; English.

XX The present sequence encodes a human cell adhesion and extracellular

CC matrix protein designated CADECM. CADECM sequences has neuroprotective,

CC cytoskeletal and anorectic activities. The CADECM polypeptides and

CC polynucleotides are useful in diagnosing, treating and preventing immune,

CC neurological, developmental, connective tissue and cell proliferative

CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon

CC cancer, obesity and tangler disease.

XX Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;

Query Match 10.4%; Score 51.4; DB 12; Length 270;

Best Local Similarity 90.2%; Pred. No. 1.7e-05;

Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTGCGAGAGAGGCTTGTGATGCGCTGAGT 60

Db 36 ATGGGACCTGATTTTGTGCTGCTGCGAGAGAGGCTTGTGATGCGCTGAGT 95

QY 61 A 61

Db 96 A 96

RESULT 3

ID ADM80826 standard; cDNA; 549 BP.

XX ADM80826;

DT 03-JUN-2004 (first entry)

DE Human CADECM-13 encoding cDNA SEQ ID NO:55.

XX human; cell adhesion and extracellular matrix protein; CADECM;
 KM neuroprotective; cytoskeletal; anorectic; immune disorder;
 KM neurological disorder; developmental disorder;
 KM connective tissue disorder; cell proliferative disorder; cancer; obesity;
 KM tangler disease; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 36..491

FT /*tag= a

/product= "CADECM-13"

FT XX WO2004015396-A2.
 PN XX
 PD XX 19-FEB-2004.
 XX
 PF 12-AUG-2003; 2003WO-US025418.
 XX
 PR 13-AUG-2002; 2002US-0403781P.
 PR 30-AUG-2002; 2002US-0407034P.
 PR 13-SEP-2002; 2002US-0410566P.
 PR 24-SEP-2002; 2002US-0413482P.
 PR 25-SEP-2002; 2002US-0413890P.
 PR 08-NOV-2002; 2002US-0424904P.
 PR 13-NOV-2002; 2002US-0426222P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JU;
 PI Wang JT, Chien D, Yang YG;
 DR WPI; 2004-191795/18.
 DR P-PSDB; ADM80784.
 XX
 PT New cell adhesion and extracellular matrix proteins, useful in
 PT diagnosing, treating and preventing immune, neurological, developmental,
 PT connective tissue and cell proliferative disorders including cancer.
 XX
 PS Claim 5; SEQ ID NO 55; 272pp; English.
 XX
 CC The present sequence encodes a human cell adhesion and extracellular
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
 CC cytostatic and anorectic activities. The CADECM polypeptides and
 CC polynucleotides are useful in diagnosing, treating and preventing immune,
 CC neurological, developmental, connective tissue and cell proliferative
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
 CC cancer, obesity and tanger disease.
 CC
 XX
 SQ Sequence 549 BP; 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
 XX

Query Match 10.4%; Score 51.4; DB 12; Length 549;
 Best Local Similarity 90.2%; Pred. No. 2.3e-05;
 Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 60
 |||
 DB 36 ATGGGACCTGATTTTATTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGCTT 95
 |||

QY 61 A 61
 |
 DB 96 A 96

RESULT 4
 ADM80827
 ID ADM80827 standard; cDNA; 623 BP.
 XX
 AC ADM80827;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human CADECM-14 encoding cDNA SEQ ID NO:56.
 XX
 XX human; cell adhesion and extracellular matrix protein; CADECM;
 XX neuroprotective; cytostatic; anorectic; immune disorder;
 KM neurological disorder; developmental disorder;
 KM connective tissue disorder; cell proliferative disorder; cancer; obesity;
 KM tanger disease; gene; ss.
 XX
 OS Homo sapiens.
 XX

Location/Qualifiers

FT CDS 36..563
 FT /*tag= a
 FT /product= "CADECM-14"

PN WO2004015396-A2.
 XX
 PD 19-FEB-2004.
 XX
 PF 12-AUG-2003; 2003WO-US025418.
 XX
 PR 13-AUG-2002; 2002US-0403781P.
 PR 30-AUG-2002; 2002US-0407034P.
 PR 13-SEP-2002; 2002US-0410566P.
 PR 24-SEP-2002; 2002US-0413482P.
 PR 25-SEP-2002; 2002US-0413890P.
 PR 08-NOV-2002; 2002US-0424904P.
 PR 13-NOV-2002; 2002US-0426222P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JU;
 PI Wang JT, Chien D, Yang YG;
 DR WPI; 2004-191795/18.
 DR P-PSDB; ADM80785.
 XX
 PT New cell adhesion and extracellular matrix proteins, useful in
 PT diagnosing, treating and preventing immune, neurological, developmental,
 PT connective tissue and cell proliferative disorders including cancer.
 XX
 PS Claim 5; SEQ ID NO 56; 272pp; English.
 XX
 CC The present sequence encodes a human cell adhesion and extracellular
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
 CC cytostatic and anorectic activities. The CADECM polypeptides and
 CC polynucleotides are useful in diagnosing, treating and preventing immune,
 CC neurological, developmental, connective tissue and cell proliferative
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
 CC cancer, obesity and Tanger disease.
 CC
 XX
 SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
 XX

Query Match 10.4%; Score 51.4; DB 12; Length 623;
 Best Local Similarity 90.2%; Pred. No. 2.4e-05;
 Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGAGT 60
 |||
 DB 36 ATGGGACCTGATTTTATTGCTGCTCTGCGAGACGCTTTGCTATGCCGTGCTT 95
 |||

QY 61 A 61
 |
 DB 96 A 96

RESULT 5
 AA250832
 ID AA250832 standard; DNA; 476 BP.
 XX
 AC AA250832;
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Rat amelogenin gene (A4).
 XX
 XX Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;
 KM chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;
 KM extracellular matrix protein; tooth enamel; enamel mineralisation;
 KM ameloblast; bone regeneration; composite cell construct; ds.
 XX

Query Match	10.4%;	Score 51.2;	DB 3;	Length 476;
Best Local Similarity	94.6%;	Pred. No. 2.5e-05;		
Matches 53;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;
Db	1	ATGGGAGACCTGATTTTGTATTCCTGCTGCGCTGGAGACACCTTTGTATGACCCGT	56	
	48	ATGGGAGACCTGATCTTTGTTTGCCTCCTCCTCGGAGACACCTTTGTATGACCCCT	103	
RESULT 6				
ADBS9026				
ID	ADBS9026	standard;	DNA;	556 BP.
XX				
AC	ADBS9026;			
XX				
DT	04-DEC-2003	(first entry)		
XX				
DE	Toxicity-related gene,	SEQ ID 4052.		
XX				
XX	Toxic; toxin; gene expression profile; hepatotoxicity; liver;			
KM	drug screening; toxicity assay; ds.			
OS	Unidentified.			
XX				
PN	WO2003064624-A2.			
XX				
XX	07-AUG-2003.			
PR	31-JAN-2003;	2003WO-US003194.		
XX				
PR	31-JAN-2002;	2002US-00060087.		
XX				
PR	15-MAR-2002;	2002US-0364045P.		
XX				
PR	15-MAR-2002;	2002US-036405P.		
XX				
PR	30-DEC-2002;	2002US-0436643P.		
XX				
PA	(GENE-) GENE LOGIC INC.			
XX				
XX	Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;			
DR	WPI; 2003-689530/65.			
XX				
PT	Predicting a toxic effect of a compound, useful in identifying toxicity			
PT	markers in liver tissues or cells for drug screening and toxicity assays,			
PT	comprises preparing gene expression profile of tissue or cells exposed to			
PT	the compound.			
XX				
PS	Claim 1; SEQ ID NO 4052; 1156pp; English.			
CC	The present invention relates to a method for predicting a toxic effect			
CC	of a compound. The method comprises preparing a gene expression profile			
CC	of a tissue or cell sample exposed to the compound, and comparing the			
CC	gene expression profile to a database comprising SEQ ID 1-4925, where			
CC	differential expression of the gene indicates at least one toxic effect.			
CC	The method is useful for predicting at least one toxic effect of a			
CC	compound, predicting hepatotoxicity or the progression of a toxic effect			
CC	of a compound, identifying an agent that modulates the onset or			
CC	progression of a toxic response, predicting the cellular pathways that a			
CC	compound modulates in a cell, and identifying an agent that modulates at			
CC	least one activity of a protein. The method and compositions of the			
CC	present invention using a database of genes having liver toxin-induced			
CC	differential expression, are useful in identifying toxicity markers in			
CC	liver tissues or cells for drug screening and toxicity assays. Note: The			
CC	sequence data for this patent did not form part of the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX				
XX	Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;			
XX				
Query Match	10.2%;	Score 50.2;	DB 10;	Length 556;
Best Local Similarity	92.9%;	Pred. No. 5.4e-05;		
Matches 52;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCCCGT 56
DB 22 ATGGGACCTGATCTGTTTGCTGCTGCTGAGACGCTTGTATGCCCGT 77

RESULT 7
ID ADB53782 standard; DNA; 556 BP.
XX ADB53782;
AC
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
XX
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KM toxicity marker; toxicity progression; drug screening;
KM primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
PN WO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003NO-US003482.
XX
PR 04-FEB-2002; 2002US-0353171P.
XX
PR 13-MAR-2002; 2002US-0363534P.
XX
PR 08-APR-2002; 2002US-0370248P.
XX
PR 10-APR-2002; 2002US-0371134P.
XX
PR 10-APR-2002; 2002US-0371135P.
XX
PR 10-APR-2002; 2002US-0371150P.
XX
PR 11-APR-2002; 2002US-0371413P.
XX
PR 19-APR-2002; 2002US-0373601P.
XX
PR 19-APR-2002; 2002US-0373602P.
XX
PR 22-APR-2002; 2002US-0374139P.
XX
PR 08-MAY-2002; 2002US-0378370P.
XX
PR 09-MAY-2002; 2002US-0378652P.
XX
PR 09-MAY-2002; 2002US-0378653P.
XX
PR 09-MAY-2002; 2002US-0378655P.
XX
PR 09-JUL-2002; 2002US-0394230P.
XX
PR 09-JUL-2002; 2002US-0394253P.
XX
PR 04-SEP-2002; 2002US-0407688P.
XX
PR 28-JAN-2003; 2003US-0442900P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Caastle A, Orr M;
PI Elashoff M;
XX
DR WPI; 2003-731472/69.
XX
PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
XX
XX Claim 44; SEQ ID NO 4324; 874bp; English.
XX
XX The present invention describes a method for determining whether a
XX compound induces a toxic effect on a tissue or cell. The method comprises
XX preparing a gene expression profile of a tissue or cell sample exposed to
XX the compound, and comparing the gene expression profile to a database
XX comprising data or information on the Tox mean and non-Tox mean value.
XX The method is useful for predicting or identifying at least one toxic
XX effect, particularly hepatotoxicity, of a test or unknown compound. The
XX genes listed in the specification are useful as diagnostic or toxicity
XX markers for the prediction or identification of the physiological state
XX of tissue or cell sample that has been exposed to a compound, or to
XX identify or predict the toxic effects of a compound or an agent. These
XX can also be used as markers for monitoring toxicity progression or for
XX drug screening. The present sequence represents a primary rat hepatocyte

CC toxicity modelling related gene sequence from the present invention.
XX
XX Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
SQ

Query Match 10.2%; Score 50.2; DB 10; Length 556;
Best Local Similarity 92.9%; Pred. No. 5,4e-05;
Matches 52; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ATGGGACCTGATTTGCTGCTGCTGAGACGCTTGTATGCCCGT 56
DB 22 ATGGGACCTGATCTGTTTGCTGCTGCTGAGACGCTTGTATGCCCGT 77

RESULT 8
ID AAL41111 standard; DNA; 722 BP.
XX AAL41111
AC AAL41111;
XX
DT 16-OCT-2002 (first entry)
XX
DE gAML related Y-chromosome DNA sequence.
XX
XX Goat embryo sexual identification technique; goat amelogenin gene; gAML;
KM sex-specific; gene; ds; Y-chromosome.
XX
OS Capra hircus.
XX
XX Key Location/Qualifiers
XX FT 35..658
XX CDS /*tag= a
XX FT /product= "Y-chromosome protein"
XX
XX TW454013-A.
XX
XX 11-SEP-2001.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX 10-NOV-1999; 99TW-00119616.
XX
XX (CHEN/) CHEN C.
XX PA (JANG/) JANG J.
XX PA (WENG/) WENG T.
XX PA (JENG/) JENG D.
XX
XX
XX Chen C, Jang J, Weng T, Jeng D;
XX
XX WPI; 2002-442016/47.
XX
XX P-PSDB; AAO22534.
XX
XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX identification, comprises high sensitivity even using single white blood
XX cell or cleavage c.
XX
XX Disclosure; Page 28; 35bp; Chinese.
XX
XX The invention relates to a goat embryo sexual identification technique
XX with high efficiency, sensitivity and repeatability. This technique
XX involves separately cloning and sequencing the coding regions and the
XX introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
XX results indicate that there are sex-specific sequences in the fifth
XX intron of the gene. The major characteristics according to the present
XX invention include high sensitivity, applicable in sex identification even
XX only using a single white blood cell or a single cleavage cell of
XX blastula; high diagnostic efficiency, capable of identifying hundreds of
XX goat embryo in 3 hours; simple operation procedures without complicated
XX steps of DNA extraction and need no additional control group intron; and
XX can be applied on different species of goats. This polynucleotide
XX sequence represents a gAML related Y-chromosome DNA sequence of the
XX invention
SQ Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;

Query Match 10.1%; Score 49.8; DB 6; Length 722;
Best Local Similarity 96.2%; Pred. No. 8e-05;
Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 53
Db 35 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 87

RESULT 9

AL41110
ID AL41110 standard; DNA; 752 BP.

AC AAL41110;

DT 16-OCT-2002 (first entry)

DE gAML related X-chromosome DNA sequence.

KM Goat embryo sexual identification technique; goat amelogenin gene; gAML;

XX sex-specific; gene; ds; X-chromosome.

OS Capra hircus.

XX Key Location/Qualifiers

FT CDS 35..658
FT /tag= a
FT /product= "X-chromosome protein"

PN TW454013-A.

PD 11-SEP-2001.

PE 10-NOV-1999; 99TW-00119616.

PR 10-NOV-1999; 99TW-00119616.

XX (CHEN/) CHEN C.

PA (JANG/) JANG J.

PA (WENG/) WENG T.

PA (JENG/) JENG D.

PI Chen C, Jang J, Weng T, Jeng D;

XX WPI: 2002-442016/47.

DR P-PSDB; AAO22534.

XX Sex-specific sequence of goat amelogenin gene, useful for embryo sexual

PT identification, comprises high sensitivity even using single white blood

PT cell or cleavage c.

XX Disclosure; Page 28; 35pp; Chinese.

XX The invention relates to a goat embryo sexual identification technique

CC with high efficiency, sensitivity and repeatability. This technique

CC involves separately cloning and sequencing the coding regions and the

CC introns of the goat amelogenin gene (gAML) on the goat chromosomes. The

CC results indicate that there are sex-specific sequences in the fifth

CC intron of the gene. The major characteristics according to the present

CC invention include high sensitivity, applicable in sex identification even

CC only using a single white blood cell or a single cleavage cell of

CC blastula; high diagnostic efficiency, capable of identifying hundreds of

CC goat embryo in 3 hours; simple operation procedures without complicated

CC steps of DNA extraction and need no additional control group intron; and

CC can be applied on different species of goats. This polynucleotide

CC sequence represents a gAML related X-chromosome DNA sequence of the

CC invention

XX Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;

XX Query Match 10.1%; Score 49.8; DB 6; Length 752;

XX Best Local Similarity 96.2%; Pred. No. 8.2e-05;

Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 53
Db 35 ATGGGAGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 87

RESULT 10

ADM80836
ID ADM80836 standard; cDNA; 318 BP.

AC ADM80836;

DT 03-JUN-2004 (first entry)

DE Human CADEC23 encoding cDNA SEQ ID NO:65.

XX human; cell adhesion and extracellular matrix protein; CADEC23;

KM neuroprotective; cytoskeletal; anorectic; immune disorder;

KM neurological disorder; developmental disorder;

KM connective tissue disorder; cell proliferative disorder; cancer; obesity;

XX Tangle disease; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 36..260
FT /tag= a
FT /product= "CADEC23"

PN MO2004015396-A2.

PD 19-FEB-2004.

PE 12-ATG-2003; 2003WO-US025418.

PR 13-AUG-2002; 2002US-0403781P.

PR 30-AUG-2002; 2002US-0407034P.

PR 13-SEP-2002; 2002US-0410566P.

PR 24-SEP-2002; 2002US-0413482P.

PR 25-SEP-2002; 2002US-0413890P.

PR 08-NOV-2002; 2002US-0424904P.

PR 13-NOV-2002; 2002US-0426222P.

XX (INCY-) INCYTE CORP.

PA Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P,

PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J,

PI Hafalia AJA, Lee SY, Jhang X, Jackson AA, Richardson TW, Blake JJ,

PI Wang JT, Chien D, Yang YG;

XX WPI: 2004-191795/18.

DR P-PSDB; ADM80794.

XX New cell adhesion and extracellular matrix proteins, useful in

PT diagnosing, treating and preventing immune, neurological, developmental,

PT connective tissue and cell proliferative disorders including cancer.

XX Claim 5; SEQ ID NO 65; 272pp; English.

XX The present sequence encodes a human cell adhesion and extracellular

CC matrix protein designated CADEC23. CADEC23 sequences has neuroprotective,

CC cytoskeletal and anorectic activities. The CADEC23 polypeptides and

CC polynucleotides are useful in diagnosing, treating and preventing immune,

CC neurological, developmental, connective tissue and cell proliferative

CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon

CC cancer, obesity and Tangle disease.

XX Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

XX Query Match 9.8%; Score 48.2; DB 12; Length 318;

XX Best Local Similarity 94.3%; Pred. No. 0.00018;

XX Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACGCTTGTATGCC 53
DB 36 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 88

RESULT 11
AAZ07020
ID AAZ07020 standard; DNA, 750 BP.

AC AAZ07020;

DT 15-NOV-1999 (first entry)

DE Amelogenin X nucleotide sequence.

XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KM multiple allelic site; apolipoprotein E; apoE; coronary artery disease;
XX Alzheimer's disease; db.

OS Unidentified.

PN WO940226-A2.

PD 12-AUG-1999.

PF 08-JAN-1999; 99WO-US000499.

PR 04-FEB-1998; 98US-00018595.

PA (PEKE) PERKIN-ELMER CORP.

PI Liyak KJ, Goodaaid F;

DR WPI; 1999-539985/45.

PT 5' nuclease amplification assay using fluorescence-quencher probes for
determination of a genotype at multiple allelic sites.

PS Disclosure; Fig 10; 95pp; English.

XX The present invention describes first and second sets of fluorescer-
quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apolipoprotein E genotype. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represents the nucleotide sequence for amelogenin X, which is
CC used in the exemplification of the present invention

XX Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Query Match 9.8%; Score 48.2; DB 2; Length 750;
Best Local Similarity 94.3%; Pred. No. 0.00026;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACGCTTGTATGCC 53
DB 69 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 121

RESULT 12
AAZ07018
ID AAZ07018 standard; DNA, 793 BP.

AC AAZ07018;

DT 15-NOV-1999 (first entry)

DE Amelogenin X nucleotide sequence.

XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
KM multiple allelic site; apolipoprotein E; apoE; coronary artery disease;
XX Alzheimer's disease; db.

OS Unidentified.

PN WO940226-A2.

PD 12-AUG-1999.

PF 08-JAN-1999; 99WO-US000499.

PR 04-FEB-1998; 98US-00018595.

PA (PEKE) PERKIN-ELMER CORP.

PI Liyak KJ, Goodaaid F;

DR WPI; 1999-539985/45.

PT 5' nuclease amplification assay using fluorescence-quencher probes for
determination of a genotype at multiple allelic sites.

PS Disclosure; Fig 8A; 95pp; English.

XX The present invention describes first and second sets of fluorescer-
quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apolipoprotein E genotype. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represents the nucleotide sequence for amelogenin X, which is
CC used in the exemplification of the present invention

XX Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 9.8%; Score 48.2; DB 2; Length 793;
Best Local Similarity 94.3%; Pred. No. 0.00026;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACGCTTGTATGCC 53
DB 69 ATGGGACCTGATTTATTTGCTGCTGCTGAGACGCTTGTGCAATGCC 121

RESULT 13
AAZ07019
ID AAZ07019 standard; DNA, 802 BP.

AC AA207019;
XX 15-NOV-1999 (first entry)
XX Amelogenin Y nucleotide sequence.
DE
XX Amelogenin X; Amelogenin Y; genotype: 5' nuclease amplification;
KM multiple allelic site; apolipoprotein E, apoE; coronary artery disease;
KM Alzheimer's disease; 88.
XX
XX unidentified.
OS
XX WO9940226-A2.
PN
XX 12-AUG-1999.
PD
XX 08-JAN-1999; 99WO-US000499.
PF
XX 04-FEB-1998; 98US-00018595.
PR
XX (PEKE) PERKIN-ELMER CORP.
PA
XX Liyak KJ, Goodaaid F;
PI
XX WPI; 1999-539985/45.
DR
XX 5' nuclease amplification assay using fluorescence-quencher probes for
PT determination of a genotype at multiple allelic sites.
PT
XX Disclosure; Fig 8B; 95pp; English.
PS
XX The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apolipoprotein E genotype. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represents the nucleotide sequence for amelogenin Y, which is
CC used in the exemplification of the present invention
XX
XX Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;
SQ
Query Match 9.5%; Score 46.6; DB 2; Length 802;
Best Local Similarity 92.5%; Pred. No. 0.00083;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTCCTCGGAGACACCTTTGCTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTCCTCGGAGACACCTTTGCTATGCC 121
RESULT 14
ADQ22977
ID ADQ22977 standard; DNA; 852 BP.
XX
AC ADQ22977;
XX
XX 26-AUG-2004 (first entry)
DT Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.
XX
DE

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM de.
KW
XX Homo sapiens.
OS
XX WO2004048938-A2.
PN
XX 10-JUN-2004.
PD
XX 26-NOV-2003; 2003WO-US038193.
PF
XX 26-NOV-2002; 2002US-0429739P.
PR
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX Aziz N, Ginsburg WM, Zlotnik A;
PI
XX WPI; 2004-441208/41.
DR
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
PT
XX Example 2; SEQ ID NO 5797; 210pp; English.
PS
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;
SQ
Query Match 9.5%; Score 46.6; DB 12; Length 852;
Best Local Similarity 92.5%; Pred. No. 0.00085;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTCCTCGGAGACACCTTTGCTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTCCTCGGAGACACCTTTGCTATGCC 121
RESULT 15
ABN28389
ID ABN28389 standard; DNA; 65 BP.
XX
AC ABN28389;
XX
XX 15-JUL-2002 (first entry)
DT Rat spliced transcript detection oligonucleotide SEQ ID NO:1137.
XX
DE Human; mouse; rat; splice transcript; detection; RNA transcript;
KM splice variant; transcriptome; oligonucleotide library; 88.
KW
XX Rattus norvegicus.
OS
XX WO200210449-A2.
PN
XX 07-FEB-2002.
PD
XX 20-JUL-2001; 2001WO-IB001903.
PF
XX

Wed Feb 9 09:28:09 2005

```

PR 28-JUL-2000; 2000US-0221607P.
PR 02-MAY-2001; 2001US-0287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Messerman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of a
PT genome, useful for detecting tissue-, pathology-, and developmental-
PT specific genes.
XX
PS Example 1; SEQ ID NO 1137; 47bp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the (sub-)
CC transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises several
CC oligonucleotides, each capable of hybridizing selectively to a set of
CC messenger RNAs transcribed from a given transcription unit of the genome,
CC which encodes one or more messenger RNA splice variants. The
CC oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a particular
CC biological or pathological state, and so allowing the detection of tissue
CC - and pathology-specific genes such as those genes only expressed in
CC specific tissue under a specific pathological condition; to detect
CC developmental specific genes; and to detect RNA transcripts and splice
CC variants of a transcriptome of a patient suffering from a particular
CC disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC rats, humans and mice, which are used in the exemplification of the
CC present invention. N.B. The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 BP; 8 A; 23 C; 14 G; 20 T; 0 U; 0 Other;

```

```

Query Match          9.24; Score 45.2; DB 6; Length 65;
Best Local Similarity 94.04; Pred. No. 0.00079;
Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 7 ACCCTGATTTTGTGCTGCTCTCGAGACGCTTGTCTATGCCCCGT 56
   |||||
DB 1 ACCCTGATTTTGTGCTGCTCTCGAGACGCTTGTCTATGCCCCCT 50

```

Search completed: February 8, 2005, 13:57:21
 Job time : 656.659 sec

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 13:03:55 ; Search time 132.88 Seconds
(without alignments)
2631.757 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492
Sequence: 1 atggggagcctggttctgtc.....aactctctctgctccca 492

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	10.4	476	4 US-09-744-128-3	Sequence 3, Appl1
2	48.2	9.8	500	2 US-09-018-595B-3	Sequence 3, Appl1
3	48.2	9.8	500	3 US-09-324-709A-3	Sequence 3, Appl1
4	48.2	9.8	793	2 US-09-018-595B-1	Sequence 1, Appl1
5	48.2	9.8	793	3 US-09-324-709A-1	Sequence 1, Appl1
6	46.6	9.5	802	2 US-09-018-595B-2	Sequence 2, Appl1
7	46.6	9.5	802	3 US-09-324-709A-2	Sequence 2, Appl1
8	41.6	8.5	1141	4 US-09-806-708B-22	Sequence 22, Appl1
9	35.8	7.3	374	4 US-08-956-171E-4303	Sequence 4303, Ap
10	35.8	7.3	374	4 US-08-781-986A-4303	Sequence 4303, Ap
11	35.8	7.3	742	4 US-08-956-171E-3662	Sequence 3662, Ap
12	35.8	7.3	742	4 US-08-781-986A-3662	Sequence 3662, Ap
13	35.8	7.3	10758	4 US-08-956-171E-221	Sequence 221, App
14	35.8	7.3	10758	4 US-08-781-986A-221	Sequence 221, App
15	35	7.1	370	4 US-09-621-976-399	Sequence 399, App
16	34.2	6.9	832	4 US-09-621-976-2813	Sequence 2813, Ap
17	33.8	6.9	832	4 US-09-621-976-2813	Sequence 2813, Ap
18	33.8	6.9	1317	4 US-09-248-796A-6415	Sequence 6415, Ap
19	33.6	6.8	474	4 US-09-621-976-18033	Sequence 18033, A
20	33.4	6.8	399	4 US-09-621-976-8976	Sequence 8976, Ap
21	33.4	6.8	16063	3 US-09-801-052-3	Sequence 3, Appl1
22	33.4	6.8	16063	4 US-10-020-121-3	Sequence 3, Appl1
23	33.2	6.7	414	4 US-09-710-279-529	Sequence 529, App
24	33.2	6.7	2896	4 US-09-710-279-4383	Sequence 4383, Ap
25	33.2	6.7	4179	4 US-09-710-279-3949	Sequence 3949, Ap
26	33.2	6.7	81001	4 US-09-750-580-1	Sequence 1, Appl1
27	33	6.7	1410	4 US-09-661-451A-13	Sequence 13, Appl1

ALIGNMENTS

28	32.8	6.7	84495	3 US-09-797-906-3	Sequence 3, Appl1
29	32	6.5	202	4 US-09-513-999C-16073	Sequence 16073, A
30	32	6.5	372	3 US-08-991-789A-26	Sequence 26, Appl
31	32	6.5	372	3 US-09-062-451-26	Sequence 26, Appl
32	32	6.5	372	4 US-09-598-326-26	Sequence 26, Appl
33	32	6.5	372	4 US-09-289-198-26	Sequence 26, Appl
34	32	6.5	372	4 US-09-429-755-26	Sequence 26, Appl
35	32	6.5	2291	4 US-09-023-655-698	Sequence 698, Appl
36	32	6.5	161652	4 US-09-497-855A-40	Sequence 40, Appl
37	31.8	6.5	7705	2 US-08-687-080-115	Sequence 115, App
38	31.6	6.4	443	4 US-09-702-705-1339	Sequence 1339, App
39	31.6	6.4	443	4 US-09-736-457-1339	Sequence 1339, App
40	31.6	6.4	443	4 US-09-614-124B-1339	Sequence 1339, App
41	31.6	6.4	443	4 US-09-671-325-1339	Sequence 1339, App
42	31.6	6.4	443	4 US-09-658-824-1339	Sequence 1339, App
43	31.6	6.4	444	4 US-09-702-705-1460	Sequence 1460, App
44	31.6	6.4	444	4 US-09-736-457-1460	Sequence 1460, App
45	31.6	6.4	444	4 US-09-614-124B-1460	Sequence 1460, App

RESULT 1
US-09-744-128-3
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Vels et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentln 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3
Query Match 10.4%; Score 51.2; DB 4; Length 476;
Best Local Similarity 94.6%; Pred. No. 1.1e-06;
Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGAGCCTGATTTGTTGCTGCTGCTGGAGACCTTGTATGCCGT 56
Db 48 ATGGGAGCCTGATTTGTTGCTGCTGCTGGAGACCTTGTATGCCGT 103
RESULT 2
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF INVENTIONS: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA

ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-018-595B-3

Query Match 9.8%; Score 48.2; DB 2; Length 500;
Best Local Similarity 94.3%; Pred. No. 1,1e-05;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 121

RESULT 3
US-09-324-709A-3
Sequence 3, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
TITLE OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-324-709A-3

Query Match 9.8%; Score 48.2; DB 3; Length 500;
Best Local Similarity 94.3%; Pred. No. 1,1e-05;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 121

RESULT 4
US-09-018-595B-1
Sequence 1, Application US/09018595B
Patent No. 596223
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-1
Query Match 9.8%; Score 48.2; DB 2; Length 793;
Best Local Similarity 94.3%; Pred. No. 1,4e-05;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTCCTGGAGACGCTTGTATGCC 121

RESULT 5
US-09-324-709A-1
Sequence 1, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-1

Query Match 9.8%; Score 48.2; DB 3; Length 793;
Best Local Similarity 94.3%; Pred. No. 1.4e-05;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 121

RESULT 6
US-09-018-595B-2
Sequence 2, Application US/09018595B
Patent No. 596223
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,595B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: PELM-744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2

Query Match 9.5%; Score 46.6; DB 2; Length 802;
Best Local Similarity 92.5%; Pred. No. 4.5e-05;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 53
DB 69 ATGGGACCTGATTTGTTGCTGCTGCTGAGAGACCTTGTATGCC 121

RESULT 7
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weltz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weltz
REGISTRATION NUMBER: 38,362

```
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-324-709A-2

Query Match          9.5%; Score 46.6; DB 3; Length 802;
Best Local Similarity 92.5%; Pred. No. 4.5e-05;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCAGTGAATTTGCTGCTGCTGCGAGACAGCTTGTATGCC 53
69 ATGGGAGCAGTGAATTTGCTGCTGCTGCGAGACAGCTTGTATGCCATGCC 121

RESULT 8
US-09-806-708B-22/c
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT APPLICATION NUMBER: US/09/806,708B
CURRENT FILING DATE: 2001-04-03
PRIORITY APPLICATION NUMBER: US 60/147,133
PRIORITY FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 1141
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: Promoter
LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.T., L.A., and B.N. FAEI promoters
US-09-806-708B-22

Query Match          8.5%; Score 41.6; DB 4; Length 1141;
Best Local Similarity 7.7%; Pred. No. 0.0023;
Matches 32; Conservative 189; Mismatches 194; Indels 0; Gaps 0;

48 TATGCCGCTGAGTAAATACCCCGCATATATTCATATTCACAAGCTGGAAATTAAG 107
541 TTTNNMMWSGBVRMAGTMMWRMNNNNNTDRTYYMMWRMBRTTYYDSCKNKSMMRG 482
108 TCTGCCCCACAGTGTGAATCTTGAAGTTTAAGACAGTACAGATGATGTCCTCAA 167
481 NNWAMMMWMAANDAGAMDHWTMTGNTMMBRAMWMMMAACRATYCCNNNNNACV 422
168 TGTCTCTGTGTTTAAAGAACCTTGGAAGCTTGTATTAATAAATAATATATTCACA 227
421 WHKHKMRMTWKYMMKACNNNNBRKAMRYAMMYGSDTNTTMMWTSMBWMTVD 362
228 TGCCCTCCACCAAGTATTCAGTGAAGAGAGAGTGGGGGAGTGGCCAGGACTCGCA 287
361 YTTMRAAMNNNNNNNNRBCKTTSMWMMMDHMTHTCTYGNNTWGSAYBMAAMSMMAAGSNB 302
288 TTTTAACAGACCTCAGAGATTCGTGAGACAAATTAATCTGTAATATCATCGCCA 347
301 VTYWCMRMTYMGKTMTNNNNNNNNNNKAMRYRTVTAMCNNRYYDDAVMTBKNTYCYAYB 242
348 TCTCTAGATGAGAACTTTTAAAGGAGCCCTTGAAAGCCCTCCAGAGAAAGTCTCG 407
241 WYEMMYGKHMBWRRABHRSMMWVCKRNKTYVSWHYAMBYBKBAVAVGNNWMD 182
```

```
408 AACAGCTTAGGCAATACACAAATGCAATTTCTCTAATAACCAATTCTA 462
181 RMAHHHMCATNNNNMMWYATYHHHMKKGAATNNKTBABRDDBAHVATYWR 127

RESULT 9
US-08-956-171E-4303/c
Sequence 4303, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 4303:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4303:
US-08-956-171E-4303

Query Match          7.3%; Score 35.8; DB 4; Length 374;
Best Local Similarity 51.6%; Pred. No. 0.094;
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

288 TTTTAACAGACCTCAGAGATTCGTGAGACAAATTAATCTGTAATATCATCGCCA 347
240 TTGTAACAACTAATCAATGATCAACAAGTGGCAAGAAATGTGAAGATTAATACCC 181
348 TCTCTAGATGAGAACTTTTGAAGGAGCCCTTGAAAGCCCTCCAGAAAGTCTCG 407
180 GATTAGTTAAAGATGCGGTGACAGCGCCGACGTTTGGACGAGATTGACCGCGCTAA 121
408 AACAGCTTAGGCAATACACAAATGCAATTTCTC 446
120 ACTAGTAAGATAGCTGTAATCGTGTACCAATGTTATC 82

RESULT 10
```


APPLICATION NUMBER: US/08/781,986A
FILING DATE: 435
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 3662:
SEQUENCE CHARACTERISTICS:
LENGTH: 742 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-3662

Query Match 7.3%; Score 35.8; DB 4; Length 742;
Best Local Similarity 51.6%; Pred. No. 0.14;
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 288 TTTTACAGCACCCTGAGAGATTCTGTGAGACATTATCTGTAATATCATCGCCCA 347
DB 408 TTGTAAACAATCTAATATGATGATCAGACTGGCAGAGAAATTTGTAATATTTACCCC 467
QY 348 TCTCTAGATGAGAGAACTTTTGAAGGACCCCTTGAAGGCTCCAGAGAAAGTCTCG 407
DB 468 GATTAGTTAAAGATGACGTGTACAAAGCCGACGTTTGCAGCGATTAGCCGGCTAA 527

QY 408 AACAGCTTAGGCAATATCTCAAAAATGCCAATTTTCTC 446
DB 528 ACTAGCTAAGATAGCTGTATCTGTGTACCAATGTATTC 566

RESULT 13
US-08-956-171E-221/C
Sequence 221, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 10758 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-956-171E-221

Query Match 7.3%; Score 35.8; DB 4; Length 10758;
Best Local Similarity 51.6%; Pred. No. 0.58;
Matches 82; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 288 TTTTACAGCACCCTGAGAGATTCTGTGAGACATTATCTGTAATATCATCGCCCA 347
DB 2860 TTGTAAACAATCTAATATGATGATCAGACTGGCAGAGAAATTTGTAATATTTACCCC 2801
QY 348 TCTCTAGATGAGAGAACTTTTGAAGGACCCCTTGAAGGCTCCAGAGAAAGTCTCG 407
DB 2800 GATTAGTTAAAGATGACGTGTACAAAGCCGACGTTTGCAGCGATTAGCCGGCTAA 2741

QY 408 AACAGCTTAGGCAATATCTCAAAAATGCCAATTTTCTC 446
DB 2740 ACTAGCTAAGATAGCTGTATCTGTGTACCAATGTATTC 2702

RESULT 14
US-08-781-986A-221/C
Sequence 221, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 10758 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-221

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 16:37:21 ; Search time 662.969 Seconds
(without alignments)
4264.119 Million cell updates/sec

Title: US-10-754-437-22
Perfect score: 492
Sequence: 1 atggggagctgagcttcttgc.....aactctctctgcctccca 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	274.6	55.8	259202	18	US-10-723-860-1311 Sequence 1311, Ap
2	124.2	25.2	600	13	US-10-027-632-287050 Sequence 287050,
3	124.2	25.2	600	13	US-10-027-632-287050 Sequence 287050,
4	123	25.0	600	13	US-10-027-632-287049 Sequence 287049,
5	123	25.0	600	15	US-10-027-632-287049 Sequence 287049,
6	48.2	9.8	500	13	US-10-104-774-3 Sequence 3, Appl1
7	48.2	9.8	500	16	US-10-455-150-3 Sequence 3, Appl1
8	48.2	9.8	793	13	US-10-104-774-1 Sequence 1, Appl1
9	48.2	9.8	793	16	US-10-455-150-1 Sequence 1, Appl1
10	48.2	9.8	793	17	US-10-755-889-605 Sequence 605, App
11	46.6	9.5	802	13	US-10-104-774-2 Sequence 2, Appl1
12	46.6	9.5	802	16	US-10-455-150-2 Sequence 2, Appl1

13	46.6	9.5	852	18	US-10-723-860-5797 Sequence 5797, Ap
14	45.2	9.2	65	10	US-09-908-975-1137 Sequence 1137, Ap
15	41.4	8.4	294	18	US-10-425-115-24717 Sequence 24717, A
16	39.2	8.0	397658	9	US-09-813-320-3 Sequence 3, Appl1
17	38.4	7.8	773	14	US-10-198-846-3870 Sequence 3870, Ap
18	38.2	7.8	798	13	US-10-027-632-10730 Sequence 10730, A
19	38.2	7.8	798	13	US-10-027-632-10731 Sequence 10731, A
20	38.2	7.8	798	15	US-10-027-632-10730 Sequence 10730, A
21	38.2	7.8	798	15	US-10-027-632-10731 Sequence 10731, A
22	37.6	7.6	585	13	US-10-027-632-238025 Sequence 238025,
23	37.6	7.6	585	15	US-10-027-632-238025 Sequence 238025,
24	36.2	7.4	3673778	15	US-10-312-881-1 Sequence 1, Appl1
25	36.2	7.4	435	13	US-10-027-632-56046 Sequence 56046, A
26	36.2	7.4	435	13	US-10-027-632-57109 Sequence 57109, A
27	36.2	7.4	435	13	US-10-027-632-57925 Sequence 57925, A
28	36.2	7.4	435	15	US-10-027-632-56046 Sequence 56046, A
29	36.2	7.4	435	15	US-10-027-632-57109 Sequence 57109, A
30	36.2	7.4	435	15	US-10-027-632-57925 Sequence 57925, A
31	36.2	7.4	595	15	US-10-029-386-7878 Sequence 7878, Ap
32	36.2	7.4	911	13	US-10-027-632-104584 Sequence 104584,
33	36.2	7.4	911	15	US-10-027-632-104584 Sequence 104584,
34	36.2	7.4	1450	18	US-10-425-115-42306 Sequence 42306, A
35	36.2	7.4	166910	16	US-10-292-337-11 Sequence 11, Appl1
36	36.2	7.4	176930	17	US-10-741-601-5679 Sequence 5679, Ap
37	36	7.3	425	13	US-10-027-632-182973 Sequence 182973,
38	36	7.3	425	15	US-10-027-632-182973 Sequence 182973,
39	35.8	7.3	374	8	US-08-781-986A-4303 Sequence 4303, Ap
40	35.8	7.3	374	16	US-10-329-624-4303 Sequence 4303, Ap
41	35.8	7.3	742	8	US-08-781-986A-3662 Sequence 3662, Ap
42	35.8	7.3	742	16	US-10-329-624-3662 Sequence 3662, Ap
43	35.8	7.3	10758	8	US-08-781-986A-221 Sequence 221, App
44	35.8	7.3	10758	16	US-10-329-624-221 Sequence 221, App
45	35.6	7.2	7536	15	US-10-252-157-159 Sequence 159, App

ALIGNMENTS

RESULT 1
US-10-723-860-1311
; Sequence 1311, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natesha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1311
; LENGTH: 259202
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1311

Query Match 55.8%; Score 274.6; DB 18; Length 259202;
Best Local Similarity 79.0%; Pred. No. 2.2e-72;
Matches 394; Conservative 0; Mismatches 89; Indels 16; Gaps 5;

QY 1 ATGGGAGCTGAGTTTGTTCCTGCTGCTGAGAGAGCTTGTGATGCGGAGT 60
DB 68873 ATGGGAGCTGAGTTTGTTCCTGCTGCTGAGAGAGCTTGTGATGCGGAGT 69932
QY 61 AAAATACCTCT-----GCATATATTCATTTTCAAGAGTTGGAATTAAGCTGCCC 114
DB 69933 AAAAAGCCCTTCATATGACATGTCATTTTCAAACTTGACATTAATATCTGCTC 69992

QY	115	CACAGTGGTAACTTTAGGGTTTAAACAGTACAGATCAGATGTCTCAAAATGTCTG	174
Db	69993	ATACTGTGTAAA--TTAGGTTTAAACAGTAAAGATCAGATGTCTTATATGTCTCT	70055
QY	175	GTTGTTTAAAGAAACCTTGGAAAGACTGTTGTAATAAAAAAATAATATCCAGATGCCTCC	234
Db	70051	GGGTTTGAAGAAACACTT-CAGGACCTGTTTAAAAAGTATATATCTCAAAATCCCGTAC	70107
QY	235	ACCCAAAGCTGATTCAGTAAAGCAGAGATGGGGGAGATGCCCAGGACTCTGCATTTTAAAC	294
Db	70110	CAAAATTTCTGATTTTGTATCAGCTGGGGCCGGG-----GCCCAGGACTCTGCATTTTAAAT	70163
QY	295	AAGCACTCAGAGATTTCTGTGGAGACAATTAACCTGTAATATATCATGCCCATCTCTAG	354
Db	70164	AAGCACTCCAGAGATTTCTGTGGAGACTGTTAGCTTGTAAATATACACACCCCATCTCTAG	70223
QY	355	ATGAGGGAACCTTTTAGAAGGAGCCCTTGAAGGCTCCAGAGAAAGTGCTTCAGACAGCT	414
Db	70224	ATGAGGGAAGCTTTTGAAGGAGCCCTTGAAGGCTCCAGAGAAAGTGCTTAAACAGCT	70283
QY	415	TAGG-CAAAATCTACAAAAATGCCAATTTTCTTAAAAACCAATTTTAAAGAGTGTCCA	473
Db	70284	TTGGAACAAATTTTACAGAGATGCCAGATTGTCTAAAAACCAATTCCTCTCAGAGATTCCA	70343
QY	474	ACTCTCTTCTGCCCTGCCA 492	
Db	70344	AATCTCTTCTGCCCTGCCA 70362	

```

RESULT 2
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; Polymorphisms in the Human Genome
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287050
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287050

```

Query Match	25.2%	Score 124.2	DB 13	Length 600
Best Local Similarity	75.3%	Pred. No. 2e-27		
Matches 195	Conservative 1	Mismatches 54	Indels 9	Gaps 3
QY	1	ATGGGAGCCTGGATTTTGTGTCCTGCGCTCCCGGAGAGAGCCTTGCATAGCCCTGAGT	60	
DB	256	ATGGGAGCCTGGATTTTATTTGCTGCGCTCCCGGAGAGCCTTGCATAGCTGTGAGT	197	
	61	AAATATCCCT-----GCATATATTCATTTCAACAGCTTGGAATTAAGCTGCGC	114	
QY				
DB	196	AAACACCCCTTGACATAGTCAGTGTCCAAATTTCAACAACTTGACATTAATAATCTGCTC	137	

QY	115	CAAGATGGTAAACCTTGGGTTTAAGCACAGACAAATCAGATGCTCCAAAGTCTCT	174
Db	136	ATAGTTGGTAAA--TTAGGTTTAAAAACAGATGAGATCAGATGCTTTCATATGTCTCT	79
QY	175	GTATTTAAGAAACACTGGAGAGGTTGTTPATAAAAAAAATATATCCCGATGCTCC	234
Db	78	GGGTTTAAACAAACACTT-CAGAGCTGTGTTTTTAAAAAGTATATTCCTCAATGCGGCTAC	20
QY	235	ACCCAGACTGATTCAGTA	253
Db	19	CAAAAATCTGATTGTA	1

```

RESULT 3
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027, 632
CURRENT FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218, 006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198, 676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193, 483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185, 218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167, 363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156, 358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146, 002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287050
LENGTH: 600
TYPE: DNA
ORGANISM: Human
US-10-027-632-287050

```

	Query Match	25.2%;	Score 124.2;	DB 15;	Length 600;
	Best Local Similarity	75.3%;	Pred. No. 2e-47;		
	Matches	195;	Conservative	1;	Mismatches 54; Indels 9; Gaps 3
QY	1	ATGGGGACCTCGATTTTGTTCCTGCTCCTCGGAGACGCTTTGCTATGCCCCGTAGT	60		
Db	256	ATGGGGACCTCGATTTTATTTGCTGCTCCTCGGAGACAGTTTGGCCATGCTGTAGT	197		
QY	61	AAATACCCCT-----GCATATATTCATTTTACAAGCTTGGAAATPAAAGTCTGCC	114		
Db	196	AAAACACCCCTTGATATAGTCAGTGCATTTTACAACCTTGGACATPAAATCTGCTGC	137		
QY	115	CACAGTTGTTAACTTTAAGGTTTAAAGACGTCACAGTCAGATGTCCTCAATGTCTCT	174		
Db	136	ATAGTTGTGTAATA--TTAGGTTTAAAAAGTATGAGATCATGATGCTCTCATATGTCTCT	79		
QY	175	GTTGTTTAAAGAAACACTTGAAGAGCTGTTATATAAAAAAATATATTCCAGATGCTCC	234		
Db	78	GGGTTGAAGAAACACTT-CAGGAGCTGTTTAAAAAGTATATTCTCAATATGCCGCTAC	20		
QY	235	ACCCAGAGCTGATTCAGTA	253		
Db	19	CAAAATTTCTGATTTGGTA	1		

RESULT 4

```
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      25.0%; Score 123; DB 13; Length 600;
Best Local Similarity 75.3%; Pred. No. 4,7e-27;
Matches 195; Conservative 0; Mismatches 55; Indels 9; Gaps 3;
```

```
QY 1 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 60
DB 256 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 197
QY 61 AAATATACCCCT-----GCATATATTCATTTTCACAGCTTGAATAAAGCTGCC 114
DB 196 AAATATACCCCTTTGATAGTCAGTCATTTTCACAACTTGACATTAATAATCTGCTC 137
QY 115 CACAGTTGTAACTTAGGCTTTAAGACAGTACAGATGATGCTCAATGTCTCT 174
DB 136 ATAGTGTGTAAA--TTAGGCTTTAAGACAGTATGATGATGATGCTCTCTCT 79
QY 175 GTGTTTAAAGAAACCTTGAAAGCTGTTATATAAATAATATATCCAGATGCTCC 234
DB 78 GGGTTGAAGAAACACTT-CAGGAGCTGTTTAAAAAGTATATTCCAATGCGCTAC 20
QY 235 ACCCAAGACTGATTCAGTA 253
DB 19 CAAAAATCTGATTTGGTA 1
```

```
RESULT 5
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
```

```
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287049
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287049
```

```
Query Match      25.0%; Score 123; DB 15; Length 600;
Best Local Similarity 75.3%; Pred. No. 4,7e-27;
Matches 195; Conservative 0; Mismatches 55; Indels 9; Gaps 3;
```

```
QY 1 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 60
DB 256 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCCGTAGT 197
QY 61 AAATATACCCCT-----GCATATATTCATTTTCACAGCTTGAATAAAGCTGCC 114
DB 196 AAATATACCCCTTTGATAGTCAGTCATTTTCACAACTTGACATTAATAATCTGCTC 137
QY 115 CACAGTTGTAACTTAGGCTTTAAGACAGTACAGATGATGCTCAATGTCTCT 174
DB 136 ATAGTGTGTAAA--TTAGGCTTTAAGACAGTATGATGATGATGCTCTCTCT 79
QY 175 GTGTTTAAAGAAACCTTGAAAGCTGTTATATAAATAATATATCCAGATGCTCC 234
DB 78 GGGTTGAAGAAACACTT-CAGGAGCTGTTTAAAAAGTATATTCCAATGCGCTAC 20
QY 235 ACCCAAGACTGATTCAGTA 253
DB 19 CAAAAATCTGATTTGGTA 1
```

```
RESULT 6
US-10-104-774-3
; Sequence 3, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
; TITLE OF INVENTION: SITES
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-774-3
```

```
Query Match      9.8%; Score 48.2; DB 13; Length 500;
Best Local Similarity 94.3%; Pred. No. 0.00045;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCC 53
DB 69 ATGGGGAACCTGATTTTGTGCTGCTGCTGAGACAGCTTTGCTATGCC 121
```

RESULT 7

US-10-455-150-3
; Sequence 3, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-455-150-3
Query Match 9.8%; Score 48.2; DB 16; Length 500;
Best Local Similarity 94.3%; Pred. No. 0.00045;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121
RESULT 8
US-10-104-774-1
; Sequence 1, Application US/10104774
; Publication No. US20020164630A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
; SITES
; FILE REFERENCE: 16842-782
; CURRENT APPLICATION NUMBER: US/10/104,774
; CURRENT FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 09/018,595
; PRIOR FILING DATE: 1998-02-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 793
; TYPE: DNA

ORGANISM: Homo sapiens
US-10-104-774-1
Query Match 9.8%; Score 48.2; DB 13; Length 793;
Best Local Similarity 94.3%; Pred. No. 0.00056;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121
RESULT 9
US-10-455-150-1
; Sequence 1, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
; Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/455,150
; FILING DATE: 04-Jun-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE: 03-June-1999
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-455-150-1
Query Match 9.8%; Score 48.2; DB 16; Length 793;
Best Local Similarity 94.3%; Pred. No. 0.00056;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 53
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCTTTGCTATGCC 121
RESULT 10
US-10-755-889-605
; Sequence 605, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:

```

APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2
SEQ ID NO 605
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-605

Query Match
Best Local Similarity 9.8%; Score 48.2; DB 17; Length 793;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 53
69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 121

Db
RESULT 11
US-10-104-774-2
Sequence 2, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 802
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-2

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 13; Length 802;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY
1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 53
69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 121

Db
RESULT 12
US-10-455-150-2
Sequence 2, Application US/10455150
Publication No. US20040093302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
Walson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
```

```

COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-10-455-150-2

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 16; Length 802;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY
1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 53
69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 121

Db
RESULT 13
US-10-723-860-5797
Sequence 5797, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:
APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
FILE REFERENCE: 05882.0193.NPUS01
CURRENT APPLICATION NUMBER: US/10/723,860
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5797
LENGTH: 852
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match
Best Local Similarity 9.5%; Score 46.6; DB 18; Length 852;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY
1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 53
69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCCTTGTATGCC 121

Db
RESULT 14
```

US-09-908-975-1137
; Sequence 1137, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Ilat
; APPLICANT: FARGER, Simchon
; TITLE OF INVENTION: Oligonucleotide library for detecting RNA transcripts and splice
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/267,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1137
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-908-975-1137

Query Match 9.2%; Score 45.2; DB 10; Length 65;
Best Local Similarity 94.0%; Pred. No. 0.0014;
Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 ACTGGATTTTGTGCTGCTGCTGAGACACCTTTGCTATGCCGT 56
DB 1 ACTGGATCTGTGTTGCTGCTGCTGAGACACCTTTGCTATGCCCT 50

RESULT 15
US-10-425-115-24717
; Sequence 24717, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 24717
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_122549C.1
US-10-425-115-24717

Query Match 8.4%; Score 41.4; DB 18; Length 294;
Best Local Similarity 54.2%; Pred. No. 0.042;
Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 58 AGTAAATACCTGCATATATTCATTTACAGCTTGAAATATAAGTCTGCCAC 117
DB 112 AGTCACTAGTCTGTTACATTCATTCATGCTCAAAAGAAATGCAATCAA 171
QY 118 AGTTGTAACCTTTAGGCTTTAAGACAGTACAGATCAAGTCTCAATGCTCTG 177
DB 172 TGTTGCGGATTTCTGTGTAACCTGTATCTTTAACTGCTCAACTATCAATTT 231
QY 178 TTTAAGAAACCTTGAAGAGCTTGTATATAAAA 212
DB 232 TTTATATATAAGTCTTGTGTTAAAAAAA 266

Search completed: February 9, 2005, 06:51:20
Job time : 667.969 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 5660.18 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-22

Perfect score: 492
Sequence: 1 atggggagactgagcttctgtt.....aactctctccgcctcca 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	130.2	26.5	516	8 AQ003910	AQ003910 CIT-HSP-2
C 2	107.6	21.9	544	8 BH118642	BH118642 RPT-24-3
C 3	98.6	20.0	513	8 AZ725508	AZ725508 RPT-24-1
4	81.6	16.6	790	8 BH080441	BH080441 RPT-24-3
5	52.8	10.7	303	7 W33932	W33932 mb54402.r1
6	52.8	10.7	430	7 W40649	W40649 mc43112.r1
7	52.8	10.7	536	7 W36345	W36345 mb7212.r1
8	52.8	10.7	585	9 AY419441	AY419441 Mus muscu
9	52.8	10.7	645	6 CD773419	CD773419 AGENCOURT
10	52.8	10.7	707	6 CB056709	CB056709 NISC_j119
11	52.8	10.7	812	3 AK029358	AK029358 Mus muscu
12	52.8	10.7	843	6 CB588212	CB588212 AGENCOURT
13	52.8	10.7	859	6 CB588525	CB588525 AGENCOURT
14	52.8	10.7	865	6 CB589177	CB589177 AGENCOURT
15	52.8	10.7	886	6 CB587051	CB587051 AGENCOURT
16	52.8	10.7	889	6 CB589251	CB589251 AGENCOURT
17	52.8	10.7	891	6 CB574837	CB574837 AGENCOURT
18	52.8	10.7	913	6 CB590451	CB590451 AGENCOURT
19	52.8	10.7	919	6 CB587332	CB587332 AGENCOURT
20	52.8	10.7	928	6 CB590111	CB590111 AGENCOURT
21	52.8	10.7	945	2 BB614068	BB614068 BB614068
22	52.8	10.7	990	7 W12906	W12906 ma89P03.r1
23	52.8	10.7	1006	7 W29475	W29475 mb99f11.r1
24	52.8	10.7	1020	7 W08102	W08102 mb40g08.r1

25	51.2	10.4	217	7 R46913	R46913 Y140 Rat in
26	51.2	10.4	223	2 BB571643	BB571643 BB571643
27	51.2	10.4	449	7 R47024	R47024 Y350 Rat in
28	51.2	10.4	1031	6 CB575508	CB575508 AGENCOURT
29	50.2	10.2	373	7 R47135	R47135 Y722 Rat in
30	50.2	10.2	395	7 R47030	R47030 Y359 Rat in
31	50.2	10.2	556	7 R46934	R46934 Y169 Rat in
32	49.6	10.1	257	7 R47100	R47100 Y601 Rat in
33	49.2	10.0	303	7 R46903	R46903 Y124 Rat in
34	49.2	10.0	318	7 R47078	R47078 Y534 Rat in
35	49.2	10.0	467	7 R46947	R46947 Y186 Rat in
36	48.6	9.9	293	7 R47143	R47143 Y79 Rat inc
37	48.2	9.8	429	6 CB473334	CB473334 sm68 B03
38	48.2	9.8	570	9 AY419439	AY419439 Homo sapi
39	48.2	9.8	570	9 AY419440	AY419440 Pan trogl
40	48.2	9.8	395	7 R46955	R46955 Y195 Rat in
41	47.6	9.7	478	7 R46933	R46933 Y166 Rat in
42	47.2	9.6	273	7 R47002	R47002 Y3 Rat incl
43	46.6	9.5	286	7 R46887	R46887 Y106 Rat in
44	46.6	9.5	783	4 BG198114	BG198114 RST17499
45	46.6	9.3	355	7 R46940	R46940 Y174 Rat in

ALIGNMENTS

RESULT 1
AQ003910/c 516 bp DNA linear GSS 26-JUN-1998
CIT-HSP-2289L4.TF CIT-HSP Homo sapiens genomic clone 2289L4,
DEFINITION
genomic survey sequence.
ACCESSION
AQ003910 GI:3081561
VERSION
AQ003910.1 GI:3081561
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 516)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL
Unpublished (1998)
COMMENT
Other GSSs: CIT-HSP-2289L4.TF
Contact: Mark Adams
Department of Bukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@igf.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.igf.org/cdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES

source

1. 516
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150375"
/db_xref="taxon:9606"
/clone="2289L4"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

ORIGIN

Query Match 26.5%; Score 130.2; DB 8; Length 516;

Best Local Similarity 82.6%; Pred. No. 8e-27; Matches 161; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 299 ACCTCAGAGATTCGTGGAGACATTAATTGTAATATCATGCCCATCTAGATGG 358
 Db 516 ACCCAGAGATTCGTGGAGACTGTAGTTGTAATATCATCACCATCTAATAGATGA 457
 QY 359 AGGAACCTTTAGAGAGACCTTGAAGGCTCCAGAGAAAGTCTGAACAGCTTAGG 418
 Db 456 AGAAGCTTTGAGAGAGACCTTGAAGGCTCCAGAGAAAGTCTTAACCACTTTGG 397
 QY 419 -CAATATCTACAAAATATCCCAATTTCTCTAAACCAATTTCTAAGAGTGCACATC 477
 Db 396 ACAATATCTACAGAGATCCAGATTTGTCATAAACCAATTCCTCTCAAGATTCACATC 337
 QY 478 TCTTCCTGCTCTCCA 492
 Db 336 TCTTCCTGCTCTCCA 322

RESULT 2
 BH118642/c 544 bp DNA linear GSS 19-JUN-2001
 LOCUS RPCI-24-359A2.TV RPCI-24 Mus musculus genomic clone RPCI-24-359A2,
 DEFINITION genomic survey sequence.
 ACCESSION BH118642
 VERSION BH118642.1 GI:14961134
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 544)
 Zhao S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other GSSes: RPCI-24-359A2.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 359 row: A column: 2
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..544
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-359A2"
 /cell_type="Spleen/Brain"
 /clone_lib="RPCI-24"
 /note="Vector: PTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

ORIGIN
 Query Match 21.9%; Score 107.6; DB 8; Length 544;
 Best Local Similarity 61.6%; Pred. No. 3.3e-20;

Matches 261; Conservative 0; Mismatches 149; Indels 14; Gaps 5;

QY 77 ATATTCATTTTCAAGCTTGAAGATTAAGTGTG-CCACAGTGTGTAACCTTAGGG 135
 Db 496 ATCCCATTTGAATTAATCTTGGAACTGAATATGCTCCCATAGTGGATTTTCACTG 437
 QY 136 TTTAAGACAGTACAGATCATGATGCTCCAAATGCTGTGTTTAAAGAACACTTGA 195
 Db 436 TGTAAACAGTGCAGATCTTAATATCCCAAGATCATTAAGGTGTTGAAGAAATCATTT 377
 QY 196 GAGCTGTATTAATAAAAAATATATCCAGATGCTCCACCAAC-TCATTCAAGTA 253
 Db 376 TTGAAGTTTGTAAAGAA-ATTCAGAAAGTCTCACTTAAGCTGTGATTAAGTT 321
 QY 254 GAGCAGAGTGGGGAGAGTCCAGAGACTGTGATTTTAAACAAGCACTCAGAGATTTCT 313
 Db 320 CAGCTGAAGTGGGACCAAGACATCAAGATTTTAAACAAGATTTGACCTTAGAGATCTT 261
 QY 314 GTGAGACATTAATTAATCTTGAATATATATGCTCCCATCTCTAGATGAGAACTTTGAA 373
 Db 260 CTTGACAGATTAGCTTGTAAACACACCTGCCATCTTAACAAGTATGATTTAGAA 201
 QY 374 GGGACCTTGAAGGCTCCAGAGAAAGTCTGAACAGCTTAGGCAATATACCAAAA 433
 Db 200 GGGACCCAGAAAGGCTTCCAGAG-AGTGTATCTTGTTTAGCAATATACAGGCC 143
 QY 434 TGCC---AATTTCTTAAACCCTTAAGAGTGTCAACTCTTCTGCCC 488
 Db 142 CCCCCCAAAAAAATCTTAAGCTTAATTAATTAATTAAGTCCAAATTTCTTGGCC 83
 QY 489 TCCA 492
 Db 82 TCCA 79

RESULT 3
 AZ725508/c 513 bp DNA linear GSS 24-JAN-2001
 LOCUS RPCI-24-10901.TV RPCI-24 Mus musculus genomic clone RPCI-24-10901,
 DEFINITION genomic survey sequence.
 ACCESSION AZ725508
 VERSION AZ725508.1 GI:12472231
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 513)
 Zhao S., Nierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
 page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
 Plate: 109 row: J column: 1
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..513
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"

/db xref="taxon:10090"
/clone="RPCI-24-10931"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notes="Vector: PTARBA1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the PTARBA1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

ORIGIN

Query Match 20.4%; Score 98.6; DB 8; Length 513;
Best Local Similarity 60.1%; Pred. No. 1.4e-17;
Matches 255; Conservative 0; Mismatches 156; Indels 13; Gaps 5;

QY 77 ATATTCATTTCACAGCTTGAATTAATTAAGTCTG-CGCCAGTGTGTAACCTTTAGG 135
DB 475 ATCCCAATTGAATTAATCTTGAATCTGAATATCTGCCATGTTGTAATTTTCACTG 416
QY 136 TTTTAAGACAGTACAGATGATGCTCTCAATGCTCTGTTTAAAGAACTTGGAA 195
DB 415 TGTAAACAGCTGCAAGATCTTAATCTCCCAAGTATCTTAATAGGTGTGAAGAAATCATTT 356
QY 196 GAGCTGTATTAATAAATAATATATTCAGATGCTCCACCCAGAC--TGAATTCAGTA 253
DB 355 TTGAGTTTGTAAATTAAGAA--NATCAGAAAGTCTCCACCTTGAAGCTGATTCAGTT 300
QY 254 GAGCAGAGAGTGGGGAGTGGCCAGACCTGCTGATTTTAAAGACCTCAGAGATCTT 313
DB 299 CAGCTGAAGTGGGAGCAAGACCTCAAGATTTTAAAGACCTTGAAGACCTTGAAGATCTT 240
QY 314 GTGGAACATTAATCTTGAATATCTCCGCTCTCTGATGAGAGAACTTTTGA 373
DB 239 CCTGTAAAGTATGCTTGAACACAGCTCCCTCTCTCAAGTATCTTAATTTGA 180
QY 374 GGGACCTTGAAGAGCTCCAGAGAAAGTGTGAACAGCTTGAAGCAATTAATCAAAA 433
DB 179 GAGAACTTGAAGAGCTTC-CAGAGAGTGTATCTTGTATGAAGCAATTAAGAGCC 121
QY 434 TGCC-----AATTTCTCTAAACCAATTTCTTAACAGATGTCCTCTCTGCCC 488
DB 120 CCCCCCAAAAATATCTTAAGCTTAATTAAGTTCAGAAATTTCTTCTGCCC 61
QY 489 TCCTA 492
DB 60 TCCTA 57

RESULT 4
BH080441 790 bp DNA linear GSS 18-JUL-2001
LOCUS
DEFINITION
RPCI-24-321F13.TJ RPCI-24 Mus musculus genomic clone
ACCESSION
BH080441
VERSION
BH080441.1 GI:14900038
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 790)
Zhaio,S., Nierman,W., Malek,J., Shatsman,S., Akincir,B., Levins,M.,
Teegye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Ruesell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 321 row: F column: 13
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers

FEATURES

Source

1. 790
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-321F13"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notes="Vector: PTARBA1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC library produced by Pieter de Jong. The
library was cloned in the PTARBA1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

ORIGIN

Query Match 16.4%; Score 81.6; DB 8; Length 790;
Best Local Similarity 62.4%; Pred. No. 1.5e-12;
Matches 181; Conservative 0; Mismatches 99; Indels 10; Gaps 3;

QY 211 AAAAATATNTCCAGATGCTCCACCAAGAC--TGATCAGTAAAGCAGAGTGGGG 268
DB 29 AAAAGAAATCAGAAAGCTTCACCTTGAAGTGAATCAATTCAGTGAAGTGGAG 88
QY 269 GAGTGCCAGAGCTGCTGATTTTAAAGACCTCAGAGATCTGTGAGACATTAAC 328
DB 89 CAGCACTCAAGATTTTAAAGACATTTGACCCCTGAGATCTTCTGTAAGATTAGC 148
QY 329 TTGTAAATATCATGCCCATCTCTAAGTGAAGAACTTTAAGAGGACCTTGAAGG 388
DB 149 TTGTAAACAACACTGCCCATCTCTAAGTGAATTTTGAAGGACCTTGAAGG 208
QY 389 CCTCCAGAGAAAGTGTCCAGACAGCTTGAAGCAATTAAGCAAAATGSC-----AATT 442
DB 209 CTTCAGAG--AGTGTACTTGTTTTGAAGCAATTAAGAGCCCCCCCCCAAAAA 266
QY 443 TCTCTAAACCAATTTCTAAGAGTGTCCACTCTCTGCTGCTCCCA 492
DB 267 TCTCTAAGCTTAATTAAGTCAAGATTCCAAATTTCTTCTGCTCCCA 316

RESULT 5
W33932 303 bp mRNA linear EST 11-SEP-1996
LOCUS
DEFINITION
mb54d02.r1 Soares mouse p3NM9F.5 Mus musculus cDNA clone
IMAGE:333219 5' similar to gb:U10095 Mouse amelogenin (MOUSE);
mRNA sequence.
ACCESSION
W33932
VERSION
W33932.1 GI:115837
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellendberg,K., Stepoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:214619
 Seq primer: mob.REGA+ET
 High quality sequence stop: 185.
 Location/Qualifiers
 1..303
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:35135"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NM19.5"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTATGCCCCCT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 10.7%; Score 52.8; DB 7; Length 303;
 Best Local Similarity 96.4%; Pred. No. 0.00032;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTCTCGGAGCAGCCTTGTATGCCCGT 56
 |||||
 Db 70 ATGGGACCTGATTTGTTGCTGCTCTCGGAGCAGCCTTGTATGCCCGT 125

RESULT 6
 W40649
 LOCUS
 DEFINITION
 430 bp mRNA linear EST 11-SEP-1996
 m40649.1 Soares mouse p3NM19.5 Mus musculus cDNA clone
 IMAGE:35135 5' similar to gb:M10095 Mouse amelogenin (MOUSE);
 mRNA sequence.
 W40649
 W40649.1 GI:1324982
 EST.
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 430)
 Reference Authors
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:223135

Seq primer: ETPRimer
 High quality sequence stop: 323.
 Location/Qualifiers
 1..430
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:35135"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NM19.5"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCATTTTCTTATGCCCCCT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 10.7%; Score 52.8; DB 7; Length 430;
 Best Local Similarity 96.4%; Pred. No. 0.00035;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTCTCGGAGCAGCCTTGTATGCCCGT 56
 |||||
 Db 58 ATGGGACCTGATTTGTTGCTGCTCTCGGAGCAGCCTTGTATGCCCGT 113

RESULT 7
 W36345
 LOCUS
 DEFINITION
 536 bp mRNA linear EST 11-SEP-1996
 mb72c12.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone
 IMAGE:334966 5' similar to gb:M10095 Mouse amelogenin (MOUSE);
 mRNA sequence.
 W36345
 W36345.1 GI:1318120
 EST.
 Mus musculus (house mouse)
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 536)
 Reference Authors
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:216366
 Seq primer: ETPRimer
 High quality sequence stop: 359.
 Location/Qualifiers
 1..536
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:334966"
 /dev_stage="19.5 dpc total fetus"

/lab host="DH10B (ampicillin resistant)"
 /clone.lib="Soares mouse p3NM19.5"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCGAGGAGGAGCGGCGGCACTTTTGTATGCCCT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Facina Bonaldo. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."

ORIGIN
 Query Match 10.7%; Score 52.8; DB 7; Length 536;
 Best Local Similarity 96.4%; Pred. No. 0.00037;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 56
 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 108

RESULT 8
 AY419441 585 bp DNA linear GSS 17-DEC-2003
 LOCUS Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,
 ACCESSION AY419441
 VERSION AY419441.1 GI:39775398
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 585)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejalraival, A.,
 Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B.,
 Ferriter, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios

TITLE Science 302 (5652), 1960-1963 (2003)
 JOURNAL 14671302
 PUBMED 2 (bases 1 to 585)
 REFERENCE Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejalraival, A.,
 Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B.,
 Ferriter, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES
 Source location/Qualifiers
 1..585
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>585
 /gene="AMELX"
 /locus_tag="HOM6891"

ORIGIN
 Query Match 10.7%; Score 52.8; DB 9; Length 585;
 Best Local Similarity 96.4%; Pred. No. 0.00038;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 56
 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 56

RESULT 9
 CD773419 645 bp mRNA linear EST 02-JUL-2003
 LOCUS AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone
 DEFINITION IMAGE:30501267 5', mRNA sequence.
 CD773419
 ACCESSION CD773419.1 GI:32431921
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 645)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Yoshiko Yamada, Takashi Nakamura, NIDCR
 cDNA Library Preparation: Clontech Laboratories, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: NDCM23 row: k column: 04
 High quality sequence stop: 611.

FEATURES
 Source location/Qualifiers
 1..645
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30501267"
 /lab_host="DH10B (phage-resistant)"
 /clone.lib="NIH_MGC_190"
 /note="Organ: Pooled - Molar; Vector: pDNR-LIB, Site 1:
 SfiI (ggccatcggcc); Site 2: SfiI (ggccgctcgcc);
 Non-normalized full-length enriched library 5' and 3'
 adapters were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTTATGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGGAGCGCGGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.71
 kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by
 PCR. This library was enriched for full-length clones and
 was constructed by Clontech Laboratories (Palo Alto, CA)
 Corp."

ORIGIN
 Query Match 10.7%; Score 52.8; DB 6; Length 645;
 Best Local Similarity 96.4%; Pred. No. 0.00039;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 56
 73 ATGGGACCTGATTTGTTGCTGCTGCTGAGACAGCTTGTATGCCCT 128

RESULT 10
 CB056709 707 bp mRNA linear EST 17-JAN-2003
 LOCUS NISC J11905.w1 Soares NM6P13-15 Mus musculus cDNA clone
 DEFINITION IMAGE:4848584 5', mRNA sequence.
 CB056709
 ACCESSION CB056709.1 GI:27794996
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 707)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapdb-remail.nih.gov
 CDNA Library Preparation:
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC)
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLM1079 row: N column: 9
 Seq primer: T7 primer.
 Location/Qualifiers
 1..707
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:4848584"
 /tissue_type="pituitary gland"
 /dev_stage="juvenile, 13-15 days"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="Soares NMBP13-15"
 /note="Organ: brain; Vector: pT73D-Pact; Site 1: NotI;
 Site 2: EcoRI; 1st strand cDNA was primed with a NotI-
 oligo(dT) primer
 5'-AATCGAAGATTCGCGCGCGCTGACGAGTGTCTTTTCTTTT-3';
 double-stranded cDNA was ligated to EcoRI adaptors
 5'-AATCGGACGAGG-3' and 5'-CCTCGCGCG-3' (Pharmacia),
 digested with NotI and cloned into the NotI and EcoRI
 sites of the pT73D-Pact vector. Library went through one
 round of normalization, and was constructed in the
 laboratory of M. Bento Soares (University of Iowa)."
 ORIGIN
 Query Match 10.7%; Score 52.8; DB 6; Length 707;
 Best Local Similarity 96.4%; Pred. No. 0.00039;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGGACCTGATTTTGTCTGCTGCTGCGAGCAGCTTGTATGCCCGT 56
 47 ATGGGACCTGATTTTGTCTGCTGCTGCGAGCAGCTTGTATGCCCGT 102
 RESULT 11
 AK029358
 LOCUS 812 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
 library, clone:4833404E21 product:amelogenin, full insert sequence.
 ACCESSION AK029358
 VERSION AK029358.1 GI:26325321
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 98279253
 PUBMED 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493374
 PUBMED 11042159

REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Saeki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hatada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL 20530913
 MEDLINE 11076861
 PUBMED 11076861
 4
 THE RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 THE PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 812)
 Adachi, J., Aizawa, K., Akiyama, T., Arai, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
 Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Nunataki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanishi, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-tesg@sc.riken.jp,
 URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: <http://genome.gsc.riken.jp/>
 URL: <http://fantom.gsc.riken.jp/>.
 Location/Qualifiers
 1..812
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:4833404E21"
 /db_xref="taxon:10090"
 /clone="4833404E21"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="0 day neonate"
 73..663
 /note="unnamed protein product; amelogenin (MGD) [GI:88005,
 GB|D31768, evidence: BLASTN, 99%, match=799]
 putative"
 /codon_start=1
 /protein_id="BAC26415.1"
 /db_xref="GI:26325322"
 /translation="MGTWLFACLLGAAPMLRPHPSGPGYINLSYEVLTPKMYQS
 MTRQPSGYEPMGWLHNOIIVLSQOHPPSHLQPHNHLVVRPAQDPVAPQPM

PVPGHSMPTQHHOPINPSPAOFPQOPQOAIIPPOSHQPMQPSPLHMQPLAPQ
PLPLPFSMQPLSLPLBELPLKAMPATDKRKREVD"

polya_signal
793..798
polya_site
812
/note="putative"
/note="putative"

ORIGIN

Query Match 10.7%; Score 52.8; DB 3; Length 812;
Best Local Similarity 96.4%; Pred. No. 0.00041;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 56
|||||
73 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 128
|||||

RESULT 12

CB588212 843 bp mRNA linear EST 03-APR-2003
LOCUS
DEFINITION
IMAGE:30293339 5', mRNA sequence.

ACCESSION
CB588212
VERSION
CB588212.1 GI:29506068
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 843)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM338 row: 9 column: 08
High quality sequence stop: 636.
Location/Qualifiers

FEATURES

source

1..843
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293339"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dt primed (5'-GACTAGTTTAAATGCGAGCGGCCCGCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 843;
Best Local Similarity 96.4%; Pred. No. 0.00041;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 56
|||||
40 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 95
|||||

RESULT 13

CB588525

CB588525 859 bp mRNA linear EST 03-APR-2003
LOCUS
DEFINITION
IMAGE:30289751 5', mRNA sequence.

ACCESSION
CB588525
VERSION
CB588525.1 GI:29506381
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 859)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM329 row: e column: 24
High quality sequence stop: 763.
Location/Qualifiers

FEATURES

source

1..859
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/issue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dt primed (5'-GACTAGTTTAAATGCGAGCGGCCCGCCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 859;
Best Local Similarity 96.4%; Pred. No. 0.00041;
Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 56
|||||
59 ATGGGAGCTGATTTGTTGCTGCTGCTGAGAGAGCTTTGCTATGCCCGT 114
|||||

RESULT 14

CB589177

CB589177 865 bp mRNA linear EST 03-APR-2003
LOCUS
DEFINITION
IMAGE:30290534 5', mRNA sequence.

ACCESSION
CB589177
VERSION
CB589177.1 GI:29507033
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM

REFERENCE
AUTHORS
TITLE
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 865)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT

Unpublished (1999)
 Contact: Robert Straubeberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: Agencourt Bioscience Corporation
 CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: NDAM31 row: f column: 15
 High quality sequence stop: 630.

FEATURES

Source

Location/Qualifiers
 1..865
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30295123"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /lab_host="NIH MGC 136"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 865;
 Best Local Similarity 96.4%; Pred. No. 0.00041;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCTGCGAGCAGCTTTGCTATGCCCGT 56
 |||||||
 DB 58 ATGGGACCTGGATTGTTGCTGCTCTGCGAGCAGCTTTGCTATGCCCGT 113

RESULT 15

CB587051

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB587051 886 bp mRNA linear EST 03-APR-2003
 AGENCOURT_12807074 NIH_MGC_136 Mus musculus CDNA clone
 IMAGE:30295123 5', mRNA sequence.
 CB587051
 CB587051.1 GI:29504907
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 886)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Straubeberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: Agencourt Bioscience Corporation
 CDNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: NDAM31 row: e column: 20
 High quality sequence stop: 619.

FEATURES

Source

Location/Qualifiers
 1..886
 /organism="Mus musculus"
 /mol_type="mRNA"

ORIGIN

Query Match 10.7%; Score 52.8; DB 6; Length 886;
 Best Local Similarity 96.4%; Pred. No. 0.00042;
 Matches 54; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTCTGCGAGCAGCTTTGCTATGCCCGT 56
 |||||||
 DB 84 ATGGGACCTGGATTGTTGCTGCTCTGCGAGCAGCTTTGCTATGCCCGT 139

Search completed: February 8, 2005, 22:18:21
 Job time: 5861.18 secs

/db_xref="taxon:10090"
 /clone="IMAGE:30295123"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 136"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:28:00 ; Search time 3791.46 Seconds
(without alignments)
8481.433 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680

Sequence: 1 atggggagactgagcttctgtc.....ggtactgagagacagacac 680

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_mt:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_dl:*

9: gb_pr:*

10: gb_ro:*

11: gb_gte:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220	32.4	5712	4	AB091793 Equus cab
2	184.2	27.1	6451	4	AB091789 Bos tauru
3	181.2	26.6	7425	4	AB091791 Sus scrof
4	162	23.8	6465	9	AB091781 Pan trogl
5	160.4	23.6	8810	9	AY040206 Homo sapi
6	160.4	23.6	158142	2	AL357130 Homo sapi
7	160.4	23.6	259202	2	AC002366 Human Xp2
8	145.2	21.4	5684	9	AB091785 Lemur cat
9	139.6	20.5	6442	9	AB091783 Saimiri s
10	135.4	19.9	5591	4	AB091794 Equus cab
11	129.8	19.1	6264	4	AB091790 Bos tauru
12	123.8	18.2	5562	4	AB091787 Ocolemur
13	118.4	17.4	1935	9	D83730 Homo sapien
14	95.2	14.0	7454	9	AB091784 Saimiri s
15	84	12.4	8004	9	AB091786 Lemur cat
16	81.4	12.0	2501	9	D83729 Homo sapien
17	81.4	12.0	7163	9	AB091782 Pan trogl
18	81.4	12.0	38765	9	BS000568 Pan trogl
19	81.4	12.0	177726	9	AC145770 Pan trogl

C	20	81.4	12.0	190089	9	BS000576 Pan trogl
C	21	81.4	12.0	200214	9	AC013412 Homo sapi
C	22	78.2	11.5	9384	10	AF294357 Mus muscu
C	23	78.2	11.5	95826	10	AL805974 Mouse DNA
C	24	73.4	10.8	212886	2	AC093946 Rattus no
C	25	73.4	10.8	234471	2	AC121424 Rattus no
C	26	72.8	10.7	695	11	BV089295 RPAMSE00
C	27	72.8	10.7	695	11	BV097603 RPAMSE00
C	28	69.2	10.2	6931	4	AB091792 Sus scrof
C	29	67.2	9.9	463	10	D8306352 Mus muscu
C	30	60.2	8.9	363	11	BV089294 RPAMSE00
C	31	60.2	8.9	363	11	BV097602 RPAMSE00
C	32	54.8	8.1	177654	2	AP000918 Homo sapi
C	33	51.6	7.6	142000	9	AP003110 Homo sapi
C	34	51.2	7.5	746	4	AF215890 Capra hir
C	35	51.2	7.5	752	4	AF215889 Capra hir
C	36	51.2	7.5	757	4	BOVAMLGNI1
C	37	51.2	7.5	820	4	BOVAMLGNI1
C	38	50.8	7.5	72832	6	AC074228 Arabidops
C	39	49.6	7.3	778	4	SSU43405 Sus scrofa
C	40	47.2	6.9	175630	2	AC010925 Homo sapi
C	41	46.6	6.9	296	10	S74899 ameleogenin
C	42	46.6	6.9	727	10	MUSMAMB
C	43	46.6	6.9	785	4	AB032194 Equus cab
C	44	46.6	6.9	789	4	AB032193 Equus cab
C	45	46.6	6.9	798	10	BC059090 Mus muscu

ALIGNMENTS

RESULT 1

AB091793 5712 bp DNA linear MAM 02-MAY-2003

LOCUS Equus caballus AMELX gene for amelogenin, partial cds.

DEFINITION AB091793

ACCESSION AB091793.1 GI:29126038

VERSION

KEYWORDS

SOURCE

ORGANISM

Equus caballus (horse)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

12672962

2 (bases 1 to 5712)

Iwase,M., Saita,Y. and Takahata,N.

Direct Submision

Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for

Advanced Studies(Sokenai), Department of Biosystems Science;

Shonan kokuasamura, Hayama, Kanagawa 240-0193, Japan

(E-mail:iwase@koryu.wol.soken.ac.jp, Tel:81-468-58-1571,

Fax:81-468-58-1544)

FEATURES

source

1..5712

/organism="Equus caballus"

/mol_type="genomic DNA"

/db_xref="taxon:9796"

/sex="male"

174..5712

/gene="AMELX"

/join(174..233,1578..1589)

/gene="AMELX"

/join(1590..1643,3566..3613,4993..5037,5305..5572)

/gene="AMELX"

/codon_start=1

/product="amelogenin"

/protein_id="BAC6112.1"

/db_xref="GI:29126039"

ORIGIN

/translation="MGWTLFASLGAFAIPLPSHPGPGYINSEYVLTPLKQVOS
LIRQYTSYGVPMGMLHIOIPLSQQNSNNALOPHHIIPVNSQHPVPOOPM
PLRQSHSVPTPOHOPNIPVPOOPFHPOPOPHOIPQOPPLHPIQPLPQPLP
PIFPLQPLPPLPPLPPLPPLPPLP"

Query Match 32.4%; Score 220; DB 4; Length 5712;

Best Local Similarity 72.6%; Pred. No. 6.5e-38;

Matches 421; Conservative 0; Mismatches 120; Indels 39; Gaps 9;

1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 60
1590 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1649
61 AAAATTCGATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 120
1650 ATGATG--CAAGTCAATTTGTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1702
121 TAGGGTTAAATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 180
1703 ---GTTAATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1758
181 TTTAAGAAACATTTTAAATCTTGATCAAGATTTTAAACAAACATTTCTCAAGTTG 240
1759 TTTAAGAAACATTTTAAATCTTGATCAAGATTTTAAACAAACATTTCTCAAGTTG 1808
241 CCTTCAC--CATTTTCTGATTCAGTATGCTGCGAGAGAGCTTCAGTATGCTGTAGT 299
1809 CCTTCACCAAGATTTGTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1862
300 TTTTAAACAGCA---AGAACCTGAGAAATTTGTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 355
1863 TTTTAAACAGCA---AGAACCTGAGAAATTTGTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1922
356 TCTCTAAGTGA--AAATTTTGAAGGAGCTTTGAAAGGCTTCAGTATGCTGTAGT 413
1923 TCTCTAAGTGA--AAATTTTGAAGGAGCTTTGAAAGGCTTCAGTATGCTGTAGT 1982
414 ATGAGTCTTT---AGTACTACAAATTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 470
1983 ATGAGTCTTT---AGTACTACAAATTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 2042
471 GATCTAATCTCTCTTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 528
2043 GATCTAATCTCTCTTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 2102
529 TTCAGTGTAGGATTTCTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 568
2103 CTCCTGTGTAGGATTTCTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 2142

RESULT 2
AB091789 6451 bp DNA linear MAM 02-MAY-2003
LOCUS AB091789
DEFINITION Bos taurus AMELX gene for amelogenin, partial cds.
ACCESSION AB091789
VERSION AB091789.1 GI:29126030
KEYWORDS
SOURCE
ORGANISM Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.

REFERENCE
AUTHORS Iwase, M., Saita, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
TITLE From the cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE 22608569
PUBMED 12672962
REFERENCE 2 (bases 1 to 6451)
AUTHORS Iwase, M., Saita, Y. and Takahata, N.
TITLE Direct Submission

JOURNAL

Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokuasimura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@min.koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)

FEATURES

source

1..6451
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/sex="male"
238..6451
/gene="AMELX"
join(238..340,1650..1661)
/gene="AMELX"
join(1662..1715,4173..4220,5560..5601,5665..5709,
5987..56451)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC6108.1"
/db_xref="GI:29126031"
/translation="MGWTLFASLGAFAIPLPSHPGPGYINSEYVLTPLKQVOS
LIRQYTSYGVPMGMLHIOIPLSQQNSNNALOPHHIIPVNSQHPVPOOPM
PLRQSHSVPTPOHOPNIPVPOOPFHPOPOPHOIPQOPPLHPIQPLPQPLP
PIFPLQPLPPLPPLPPLPPLPPLP"

gene

5' UTR

CDS

ORIGIN

Query Match 27.1%; Score 184.2; DB 4; Length 6451;

Best Local Similarity 67.6%; Pred. No. 4.1e-30;

Matches 398; Conservative 0; Mismatches 143; Indels 48; Gaps 8;

1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 60
1662 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1721
61 AAAATTC-----CAATTCGAATTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 108
1722 AAAACCTCTCTTCAAAAGTAAATGTCAGTTTCAAGCTTGAAGAGAGAGCTTCAGTATGCTGTAGT 1781
109 CAGATTTTATTTAGGTTAAATTTAAGGTTTAAAGAGAGAGCTTCAGTATGCTGTAGT 168
1782 CACAGT-----GGTAAAGGTTTAAAGAGAGAGCTTCAGTATGCTGTAGT 1832
169 AATGACCTGTGTTTAAAGCAATTTTAAATCTTGATCAAGATTTTAAACAAAAA 228
1833 AATGCTCTGTGTTTAAAGCAATTTTAAATCTTGATCAAGATTTTAAACAAAAA 1880
229 CATTCAGTGTGCTTCAACCAATTTGATTCAGTATGCTGCGAGAGAGCTTCAGTATGCTGTAGT 288
1881 GATTCAGATGTGCTTCTGCTCA-----AGATTCGATTCAGTATGCTGCGAGAGAGCTTCAGTATGCTGTAGT 1935
289 AGAAGCTGAAATTTTAAACAGCA---AGAACCTGAGAAATTTCTGGAACAAATTAATT 344
1936 AGAAGCTGAAATTTTAAACAGCA---AGAACCTGAGAAATTTCTGGAACAAATTAATT 1995
345 AATATCATCTCATCTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 402
1996 AATATCATCTCATCTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 2055
403 AAAAGTCTCATCTGCTT---AAGTACTACAAATTTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 459
2056 GAAAGTATTCACAGCTTGAAGCAATTTTAAAGATTTGATGATTTCTTAAATTTCA 2115
460 ATTTCACAGATTTCACTCTCTTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 519
2116 GCTTTTCAATTAAGTCAAAATTTTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 2175
520 GA-GTAAATTTCAATTTGATTTCTGCTGCTGCGAGAGAGCTTCAGTATGCTGTAGT 567
2176 GAGGTAAATTTTGTGTTAGGAATTTCTTCTTGAACCAATCTCCCTG 2224

RESULT 3
AB091791 7425 bp DNA linear MAM 02-MAY-2003
LOCUS
DEFINITION
Sus scrofa AMELX gene for amelogenin, partial cds.
AB091791
AB091791.1 GI:29126034
KEYWORDS
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE
1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
JOURNAL
MEDLINE
22608569
PUBMED
12672962
2 (bases 1 to 7425)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submision
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
Location/Qualifiers
1. .7425
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/sex="male"
1491. .7425
/gene="AMELX"
join(1491. .1548,3140. .3151)
/gene="AMELX"
join(3152. .3205,5316. .5363,6619. .6660,6717. .6761,
7033. .7425)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66110.1"
/db_xref="GI:29126035"
/translation="MGWILIFACLLGAARFMPHPHGHGYINFSYEDLYEAIRID
RFAVLTPLKWTQNMIRHPTYSYGYEPMGMLHNOIIPVVSQTPSHLOPHNHPM
VPAQOQPIPOQPMPIRGHSMPTQHNQNPRLPRAQOFPQOFPQOFPQOFPQOFP
MHPQPLPQPLPMPFMSQSLPDLPLEAMPAT"
ORIGIN
Query Match 26.6%; Score 181.2; DB 4; Length 7425;
Best Local Similarity 67.0%; Pred. No. 1.8e-29;
Matches 391; Conservative 0; Mismatches 158; Indels 35; Gaps 8;
QY 1 ATGGGACCGGATTTTGTTCCTGCTGCGGAGACCTTCAGTATGCTGAGT 60
DB 3152 ATGGGACCGGATTTTGTTCCTGCTGCGGAGACCTTCAGTATGCTGAGT 3211
QY 61 AAAATTTC-----AATTCATTTTCAAGCTTGAAATAATAATTCGCT 108
DB 3212 AAAGACCCCTTCGATCGTACATGTCATGTCATGCTTGGAATAATAATTCACC 3271
QY 109 CAGATTTTATTAAGGTTAAATTAAGGTTAAGACAGTCAAGATCTGATGTTTACA 168
DB 3272 CCAGAT-----GATACAGTTAGGATGAAAAACATGAAGATCAATTTCTCA 3322
QY 169 AATGTAGCTGTGTTAAGAAACATTTTAAATCTTGATCAACAAGATTTTAAAAA 228
DB 3323 AATGACTCAGTGTGTTAAGAAACATTTGAAACCTT-----GTTACAAAAA 3374
QY 229 CATTCAGTTGCTTCAACCAATATTCATTCAGTATGACGAGAGTGGGGGTGTG-C 286
DB 3375 GATTCACAGATGCTCTGCCAAGATTCATTCAGTATGCTGGGGGTGGGCCATGAC 3434

QY 287 TGGAACCTCTGAATTTTAAACAGCAAGACCTGAGAAATCTTGGAACAATTACTTAT 346
DB 3435 TCTGCATTTTAAACAGCACCCGAGAGAGCTCTCTTAAACAATTAATCTGTAATCAT 3494
QY 347 AATCAGTCAATCTAGATGAAATTTTGGAAAGGACCTTTGAAAGGCTCC--AGAA 404
DB 3495 CACCTGCTCTAGAGAGGAACT--TTTGAAGTACCTTTAGAGGCTCCAGAGAA 3553
QY 405 AAGTCTCAATCAGCTTTTAAAGTACTCAAAAATGACAGTTTCTTAAACCAATTTTC 464
DB 3554 TACTGAAGACGCTTTAGCAAAATAGACAAAATGACAGTTTCTTAAATCAATTTTC 3613
QY 465 TCACAAGTATCTACTCTTTTGTCTCAGAAAACCCCGGCTATCATCTTCAAG-T 523
DB 3614 TCCCAAGTCCCAAAATATCTTCTGCTCTCCAAACCCCGCTGTATTTCAAGAGAG 3673
QY 524 AAGATTCAGTGTAGAAATTTCTTCTGCACTATTTATTTGT 567
DB 3674 TAAGATTTGTGTAGAAATTTCTTCTGCACTATTTCTGT 3717
RESULT 4
AB091781 6465 bp DNA linear PRI 02-MAY-2003
LOCUS
DEFINITION
Pan troglodytes AMELX gene for amelogenin, partial cds.
AB091781
AB091781.1 GI:29126015
KEYWORDS
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE
1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
From the cover: The amelogenin loci span an ancient pseudautosomal
boundary in diverse mammalian species
Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
JOURNAL
MEDLINE
22608569
PUBMED
12672962
2 (bases 1 to 6465)
Iwase, M., Satta, Y. and Takahata, N.
Direct Submision
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies (Sokendai), Department of Biosystems Science;
Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
(E-mail: iwase@koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
Fax: 81-468-58-1544)
Location/Qualifiers
1. .6465
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/sex="male"
927. .6465
/gene="AMELX"
join(927. .982,2283. .2294)
/gene="AMELX"
join(2295. .2348,4284. .4331,5607. .5648,5740. .5781,
6052. .6465)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66101.1"
/db_xref="GI:29126016"
/translation="MGWILIFACLLGAARFMPHPHGHGYINFSYENSHSOALINVD
RFAVLTPLKWTQNMIRHPTYSYGYEPMGMLHNOIIPVVSQTPSHLOPHNHPM
VPAQOQPIPOQPMPIRGHSMPTQHNQNPRLPRAQOFPQOFPQOFPQOFPQOFP
MHPQPLPQPLPMPFMSQSLPDLPLEAMPSTDKX"
ORIGIN
Query Match 23.8%; Score 162; DB 9; Length 6465;
Best Local Similarity 65.3%; Pred. No. 2.8e-25;

Matches 385; Conservative 0; Mismatches 155; Indels 50; Gaps 8;

Qy 1 ATGGGACCTGGATTGTTGTTGCTGCTCTGGAGAGAGCTTCAAGTATGCTGTAGT 60
 Db 2295 ATGGGACCTGGATTGTTGTTGCTGCTCTGGAGAGAGCTTGTGCAATGCTGTAGT 2354

Qy 61 AAAATTC-----CAATTCGAATTCAGAGCTTGGAAATTAATAATCTGCT 108
 Db 2355 AAAACACCCCTGCATAGTCAGTGCATATTCACAAAGCTTGGACATTAATAATCTGCC- 2413

Qy 109 CAGATTTTATTTAAGGTTAAATTTAAGGTTTAAAGAGTAAAGATCTGATGTTTACA 168
 Db 2414 -----CATGTTGTTGAATTAAGGTTTAAACAGTATGATGATGATGCTTCA 2463

Qy 169 AATGTCATGTTTAAAGAAACATTTTAAATCTTATGATCAAGATTTTAAACAAAA 228
 Db 2464 TATGTCCTGGGTTGAAGAAACACTTCAGAGCTT-----GTTTAAAAAGTA 2512

Qy 229 CATTCAGTTGCTTCAACCATATTCGATTCAGTATAGCAGAGTGGGGGTGTGCTG 288
 Db 2513 TATTCCTAATGCCGCTACCAAA-----ATTGATTTGTATACAGCTGGGGGGGCC 2567

Qy 289 AGAATCTGAATTTTAAACAACA-----AGAACTCAGAAATTCCTTGAACAATTAATT 344
 Db 2568 AGAATCTGAATTTTATTAAGCAACCCAGAGATTCGTGGAACGTGTAAGCTTGAAT 2627

Qy 345 ATATCATCATCTCTGATGGA--AAATTTTGAAGGAGCTTTTAAAGGCTCCAGA 402
 Db 2628 ATCCACCCATCTCTGATGAGAGAGAGAGCTTTTGAAGGAGAGCTTGAAGGCTCCAGA 2687

Qy 403 AAAAGTCTCAATCAGTCTT---AAGTACTACAAATGCGAGTTTCTTAAACCAA 459
 Db 2688 GAAAGTCTTAAACAGCTTGGACAAATATTAAGAGATGCGAGTTTGTCTTAAACCCA 2747

Qy 460 ATTTCACAAAGTATCTAATCTCTTCTGCTGCAAAACCCCGGCTATCAATTTCTCA 519
 Db 2748 ATTTCCTCAAGATTCCAAACTCTCTGCTGCTCCACATATTCGCTTATCCCTCA 2807

Qy 520 --GAGTAAGATTCAGTGTAGGAATTCCTGCTGCAATTAATTTGT 567
 Db 2808 GGGGTAAGATTTTGTGTTAGGAATTCACATTTTGAAGCAATTCCTGT 2857

RESULT 5
 AY040206 8810 bp DNA linear PRI 10-APR-2002
 LOCUS Homo sapiens amelogenin precursor (AMELX) gene, complete cds.
 ACCESSION AY040206 GI:15028582
 VERSION AY040206.1 GI:15028582
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Hart,P.S., Hart,T.C., Simmer,J.P. and Wright,J.T.
 TITLE A nomenclature for X-linked amelogenesis imperfecta
 JOURNAL Arch. Oral Biol. 47 (4), 255-260 (2002)
 MEDLINE 21920287
 PUBMED 11922868
 REFERENCE
 2 (bases 1 to 8810)
 AUTHORS Hart,S., Hart,T.C. and Wright,T.
 TITLE Direct Submission
 JOURNAL Submitted (14-JUN-2001) Human Genetics, University of Pittsburgh,
 3550 Terrace St., 572A Scaife Hall, Pittsburgh, PA 15090, USA
 FEATURES
 source
 1. 8810
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp22.3-p22.1"
 1. 8791

promoter
 1. .1416
 /gene="AMELX"

TATA_signal
 1386. .1391
 /gene="AMELX"

mRNA
 join(1417. .1472,2781. .2846,4783. .4830,6240. .6281,
 6552. .6977,8611. .8791)
 /gene="AMELX"

exon
 1417. .1472
 /product="amelogenin precursor"
 /gene="AMELX"

exon
 2781. .2846
 /number=1
 /gene="AMELX"

CDS
 join(2793. .2846,4783. .4830,6240. .6281,6552. .6977,
 8611. .8616)
 /gene="AMELX"

sig_peptide
 2793. .2840
 /gene="AMELX"

exon
 4783. .4830
 /gene="AMELX"

exon
 6107. .6148
 /gene="AMELX"

exon
 6240. .6281
 /gene="AMELX"

exon
 6552. .6977
 /number=5
 /gene="AMELX"

exon
 8611. .8791
 /number=6
 /gene="AMELX"

exon
 8785. .8791
 /number=7
 /gene="AMELX"

polyA_signal
 8785. .8791
 /gene="AMELX"

ORIGIN
 Query Match 23.6%; Score 160.4; DB 9; Length 8810;
 Best Local Similarity 65.1%; Pred. No. 6e-25;
 Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;

Qy 1 ATGGGACCTGGATTGTTGTTGCTGCTCTGGAGAGAGCTTCAAGTATGCTGTAGT 60
 Db 2793 ATGGGACCTGGATTGTTGTTGCTGCTCTGGAGAGAGCTTGTGCAATGCTGTAGT 2852

Qy 61 AAAATTC-----CAATTCGAATTCAGAGCTTGGAAATTAATAATCTGCT 108
 Db 2853 AAAACACCCCTGCATAGTCAGTGCATATTCACAAAGCTTGGACATTAATAATCTGCC- 2910

Qy 109 CAGATTTTATTTAAGGTTAAATTTAAGGTTTAAAGAGTAAAGATCTGATGTTTACA 168
 Db 2911 -----TATAGTGTGTTGAATTAAGGTTTAAACAGTATGATGATGCTTCA 2961

Qy 169 AATGTCATGTTTAAAGAAACATTTTAAATCTTATGATCAAGATTTTAAACAAAA 228
 Db 2962 TATGTCCTGGGTTGAAGAAACACTTCAGAGCTT-----GTTTAAAAAGTA 3010

Qy 229 CATTCAGTTGCTTCAACCATATTCGATTCAGTATAGCAGAGTGGGGGTGTGCTG 288
 Db 3011 TATTCCTAATGCCGCTACCAAA-----ATTGATTTGTATACAGCTGGGGGGGCC 3065

Qy 289 AGAATCTGAATTTTAAACAACA-----AGAACTCAGAAATTCCTTGAACAATTAATT 344

Db 3066 AGACCTGCTGATTTTAAAGACCCAGGAGATTCTGTGGAACGTGTGTAAT 3125
Qy 345 ATATCTCTCATCTCTTGATGGA--AAATTTTGAAGGACCTTTGAAAGGCTCCAGA 402
Db 3126 ATCCACCCCATCTCTTGATGAGAGAGCTTTTGAGAGGACCTTTGAAAGGCTCCAGA 3185
Qy 403 AAAAGTCTCATGCTCTT---AGTACTACAAAATGCAAGTTTCTCTTAACCA 459
Db 3186 GAAAGTCTTAAACAGCTTTGACAAATATTTACAGAGATGCGAGTTTGTCTAAACCA 3245
Qy 460 ATTCTCACAAGATATCTACTCTCTTCTGCTCAGAAACCCCGGCTATCATCTTCA 519
Db 3246 ATTCCTCTCAAGATCCAAATCTCTCTGCTCCGCCACATATGCTGTCTACCCCTCA 3305
Qy 520 --GAGTAAGATTTCAGTGTAGGAATTTCTCTCTGACTATTTATTTGT 567
Db 3306 GGGGGTAAGATTTTGTGTAGGAATTCACATTTTGGAGCCACATTCCTGT 3355

RESULT 6
AL357130/c 158142 bp DNA linear HTG 10-JUN-2001
LOCUS Homo sapiens chromosome X clone RP13-169E15, 4 unordered pieces.
DEFINITION AL357130.3 GI:9863814
ACCESSION
VERSION HTG; HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
McLay, K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requester: clonerequest@sanger.ac.uk
On Aug 21, 2000 this sequence version replaced gi:9214076.

COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: B169E15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 155448 bases at least Q40
Consensus quality: 157038 bases at least Q20
Insert size: 157842; sum-of-contigs
Insert size: 160705; 33.4% error; agarose-fp
Quality coverage: 4.56x in Q20 bases; sum-of-contigs Quality
coverage: 4.56x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 68897: contig of 68897 bp in length
* 68898 68997: gap of 100 bp
* 68998 122842: contig of 53845 bp in length
* 122843 122943: gap of 100 bp
* 122943 125584: contig of 6642 bp in length
* 125585 129685: gap of 100 bp
* 129685 158142: contig of 28456 bp in length.
Location/Qualifiers
1. 158142

FEATURES
Source

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-169E15"
/clone_id="RP13-169E15"
1. 68897
/note="assembly; fragment: 01108
fragment_chain: 1
clone_end: T7
vector_side: left"
68998 - 122842
/note="assembly; fragment: 00101
fragment_chain: 1"
122943 - 129584
/note="assembly; fragment: 00361
fragment_chain: 1"
129685 - 158142
/note="assembly; fragment: 01285
fragment_chain: 1
clone_end: Sp6
vector_side: right"

ORIGIN

Query Match 23.6%; Score 160.4; DB 2; Length 158142;
Best Local Similarity 65.1%; Pred. No. 4e-25;
Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;

Qy 1 ATGGGACCTGGAATTTGTTGGCTGCTGCTGGAGCAGCTTCAGTATGCTGTGAGT 60
Db 44325 ATGGGACCTGGAATTTGTTGGCTGCTGCTGGAGCAGCTTCAGTATGCTGTGAGT 44266
Qy 61 AAAATTC-----CAATTCGAATTTGCAAGCTGGAAATTAATTCGCT 108
Db 44265 AAAAACCCTTCGATAGTCAGTTCGAATTTGCAAACTTGACATATAAATTCGCT-- 44208
Qy 109 CAGATTTTATTTAGGTTAAATTTAAGGTTTAAAGCTAACAAGATCTGATGTTTACA 168
Db 44207 -----TCATAGTTGTTGTAATTTGATTTAAACGATAGATGATGTTCTTCA 44157
Qy 169 AATGACCTGTTGTTAAGAAACATTTTAAATCTTGATCACAAGATTTTAAACAAA 228
Db 44156 TATGTCCTGCTGTTGAAGAAACCTTCAGAGCTT-----GTTTAAAGATTA 44108
Qy 229 CATTCTCAGTTCCTTACCACCATATTTCTGATTCAGTATGACAGAGTGGGGCTGCTG 288
Db 44107 TATTCTCAATTTGCGGTACCAAAA-----ATTCTGATTTGTTGACGCTGGGGCGGGGCC 44053
Qy 289 AGAATCTGAATTTTAAACAGCA-----AGAACTCAGAAATCTTGAAACAAATTA 344
Db 44052 AGAATCTGAATTTTAAACAGCA-----AGAACTCAGAAATCTTGAAACAAATTA 43993
Qy 345 ATATCACTCATCTCAGATGGA--AAATTTTGAAGGACCTTTGAAAGGCTCCAGA 402
Db 43992 ATATCACTCATCTCAGATGGA--AAATTTTGAAGGACCTTTGAAAGGCTCCAGA 43933
Qy 403 AAAAGTCTCATGCTCTT---AGTACTACAAAATGCAAGTTTCTCTTAACCA 459
Db 43932 GAAAGTCTTAAACAGCTTTGACAAATATTTACAGAGATGCGAGTTTGTCTAAACCA 43873
Qy 460 ATTCTCACAAGATATCTACTCTCTTCTGCTCAGAAACCCCGGCTATCATCTTCA 519
Db 43872 ATTCTCTCAAGATCCAAATCTCTCTGCTCCACATATGCTGCTTTTACCCCTCA 43813
Qy 520 --GAGTAAGATTTCAGTGTAGGAATTTCTCTGACTATTTATTTGT 567
Db 43812 GGGGGTAAGATTTTGTGTAGGAATTCACATTTTGGAGCCACATTCCTGT 43763

RESULT 7
AC002366 259202 bp DNA linear PRI 11-JUN-1998
LOCUS Human Xp22 BAC CT-285115 (from Caltech/Research Genetics) , PAC
DEFINITION


```

repeat_region      /rpt_family="MER20"
                    complement(61290..61384)
                    /rpt_family="L1PB1"
repeat_region      61683..61811
                    /rpt_family="MER20"
                    /rpt_family="L1PB7"
repeat_region      61907..62113
                    /rpt_family="L1MB7"
                    complement(63588..63969)
                    /rpt_family="L1MA10"
repeat_region      64800..65073
                    /rpt_family="Alu"
                    complement(67157..67209)
                    /rpt_family="MIR"
repeat_region      67637..75845
                    /gene="AMELX"
                    /gene="AMELX"
                    /gene="AMELX"
                    /note="similar to bovine amelogenin promoter"
                    join(68495..68552,69861..69926,71063..71910,73320..73361,
                    73632..74057,75691..75845)
                    /gene="AMELX"
                    /product="amelogenin"
                    join(69873..69926,71863..71910,73320..73361,73632..74057,
                    75691..75696)
                    /gene="AMELX"
                    /note="similar to amelogenin encoded by GenBank Accession
                    Number M86932"
                    /codon_start=1
                    /product="amelogenin"
                    /protein_id="AAC21581.1"
                    /db_xref="GI:2739350"
                    /translation="MGTWITFACILGAFAMPLPRPHGPIYNSYEVLTPLKWTOS
                    IRPRYSYGRPMGMLHQITIVLSQDHPRTTLQPHNIPVPRQOPVTPQOPMP
                    VPOQHSMTPIQHQPMLPRPAQOPVOPQOPVOPQOPVOPQOPVOPQOPVOPQOPVOPQOPVOP
                    MPMQPLPRLPDLTLEAMPSTDKTKKEEVD"
                    complement(77161..77331)
                    /rpt_family="MER3"
repeat_region      /rpt_family="MER3"
                    complement(79184..79473)
                    /rpt_family="Alu"
repeat_region      /rpt_family="Alu"
                    complement(79881..80169)
                    /rpt_family="Alu"
repeat_region      80196..80253
                    /rpt_family="MER20"
                    /rpt_family="L1PB3"
repeat_region      80384..80459
                    complement(81544..81633)
                    /rpt_family="MIR"
repeat_region      84789..84957
                    /rpt_family="L1MA2"
repeat_region      84978..85252
                    /rpt_family="Alu"
repeat_region      85294..85692

Query Match      23.6% Score 160.4; DB 9; Length 259202;
Best Local Similarity 65.1%; Pred. NO. 3.7e-25;
Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;

```

```

QY      229 CATTCTCAGTTGCTTACCCCATATTCGATTCAATATGAGGAGTGGGGTGTGCTG 288
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70091 TATTCCTAATGCGCTTACCAAAA-----ATTCCTATTGGTACCTGGGGGGGGCC 70145
QY      289 AGAAGCTGAAATTTTAAACAGCA---AGAACTCAGAAATCTTGAAACATTAATT 344
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70146 AGAGCTCTCATTTTATAAGACACCCAGAGATTCGTGGAACTGTTGTAAT 70205
QY      345 ATPATCTCATCTCTAGATGCA--AAATTTTGAAGGAGCTTTGAAAGCCTCAGA 402
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70206 ATCACCACTCATCTCTAGATGAGGAGAGCTTTTGAAGGAGCCTTGAAAGCTCAGA 70265
QY      403 AAAAGTCTCAATCAGTCTT---AGTACTCAAAAATGCCAGTTTCTTAAACCA 459
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70266 GAAAGTCTTACACACCTTTGACAAATATTTACAGAGAGCCAGTTTGTCTTAAACCA 70325
QY      460 ATTTCTCAGAATATTAATCTCTTCTGCTCAGAAAACCCCGGCTATCAATCTTCA 519
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70326 ATTCCTCAGAATTTCAATATCTTCTGCTCAGCAATATTCGCTTACCCCTCA 70385
QY      520 --GAGTAAGATTCACTGTTAGGAATTCCTTCTGTAATTAATTGT 567
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      70386 GGGGTAAGATTTTGTGTAGGAATCACTTTTGAGCCACATTCCTGT 70435

```

```

RESULT 8
AB091785      5684 bp      DNA      linear      PRI 02-MAY-2003
LOCUS      Lemur catla AMELX gene for amelogenin, partial cds.
DEFINITION      AB091785
ACCESSION      AB091785.1 GI:29126023
VERSION
KEYWORDS
SOURCE
ORGANISM
Lemur catla (ring-tailed lemur)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsithini; Lemnidae; Lemur.
REFERENCE
AUTHORS      Iwase,M., Saita,Y., Hirai,Y., Hirai,H., Imai,H. and Takahata,N.
TITLE      Boundary in diverse mammalian species
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE      22608569
PUBMED      12672962
REFERENCES
2 (bases 1 to 5684)
Iwase,M., Saita,Y. and Takahata,N.
Direct Substition
Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
Advanced Studies(Soken dai), Department of Biosystems Science;
Shonan Kokusaijima, Hayama, Kanagawa 240-0193, Japan
(E-mail:iwase@koryu.wol.soken.ac.jp, Tel:81-468-58-1571,
Fax:81-468-58-1544)
FEATURES
source
1..5684
/organism="Lemur catla"
/mol_type="genomic DNA"
/db_xref="taxon:9447"
/sex="male"
100..5684
/gene="AMELX"
join(100..156,1503..1514)
/gene="AMELX"
CDS
5' UTR
100..5684
/gene="AMELX"
join(1515..1560,3465..3512,4804..4845,4937..4981,
5229..5684)
/gene="AMELX"
/codon_start=1
/product="amelogenin"
/db_xref="GI:29126024"
/translation="MGTWITFACILGAFAMPLPRPHGPIYNSYENSHQALISY
GTELVTPLKWTOSMIRPRYSYGRPMGMLHQITIVLSQDHPRTTLQPHNIPV
VOAQOPVVPQOPMPVPOQHSMTPIQHQPMLPRPAQOPVOPQOPVOPQOPVOPQOPVOP
MPMQPLPRLPDLTLEAMPSTDKTKKEEVD"
ORIGIN

```



```

VERSION      AB091794.1  GI:29126040
KEYWORDS
SOURCE
ORGANISM      Equus caballus (horse)
               Equus caballus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE
AUTHORS      1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
TITLE        From the cover: The amelogenin loci span an ancient pseudautosomal
               boundary in diverse mammalian species
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE      22608569
PUBMED       12672962
REFERENCE
TITLE        2 (bases 1 to 5591)
AUTHORS      Iwase, M., Satta, Y. and Takahata, N.
JOURNAL      Direct Submission
               Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
               Advanced Studies (Sokendai), Department of Biosystems Science;
               Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
               (E-mail: iwase@n.koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
               Fax: 81-468-58-1544)
FEATURES
source       Location/Qualifiers
             1..5591
             /organism="Equus caballus"
             /mol_type="genomic DNA"
             /db_xref="taxon:9796"
             /sex="male"
             130..5591
             /gene="AMELY"
             join(130..160,1483..1494)
             /gene="AMELY"
             join(1495..1548,3470..3517,4907..4915,5235..5591)
             /gene="AMELY"
             /codon_start=1
             /product="amelogenin"
             /protein_id="BAC66113.1"
             /db_xref="GI:29126041"
             /translation="MGTWILFACLVGTAIAMP.PHPRGHAYINFSYVLTPLKX
             PMSGLHQLIPVLTQONPSNHLQPHNHIPVPAQOPVVSQOPALPLEGSHMPSQ
             HHOPNLPVPQOPFOPQOPHOPLOPQPPMHPIOPLPQAPLPPIPLAASGPHAS"
ORIGIN
Query Match      19.9%; Score 135.4; DB 4; Length 5591;
Best local Similarity 62.4%; Pred. No. 1.8e-19;
Matches 360; Conservative 0; Mismatches 181; Indels 36; Gaps 8;
QY      1 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGACCTTCAGTATCGCTGAGT 60
DB      1495 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGACCTTCAGTATCGCTGAGT 1554
QY      61 AAAATTTCCATTTCCATTTTCACAGCTTGGAAATTAATCTGCTCAGATTTTATT 120
DB      1555 ATTAAT--CAGCATCTATTTCACAGCTTGAATGCAAAAATGGCTCA-----T 1603
QY      121 TAGGGTTAAATTTAAGGGTTAAGACAGTACAGATCTGATCTTCAAAATGAGCTG 180
DB      1604 TGTGGTAAATTTTGGGGTTAATCAATCAAGATCATGTATCTTCAATGTCTCAG 1663
QY      181 TTTAAGAAACATTTTAAATCTGTATCACAAGATTTTAAACAAAACATTTCTCAGTT 240
DB      1664 TTTAAGATTTAATTTGAAGAGT-----AGTTATTAAGAAATAGATTTCTCGAG 1713
QY      241 CCTCAGCCATATTCGATTCAGTATAGCCAGAGTGGGGGGTGTCTGAGAACTCGAAT 300
DB      1714 ACTCTGCCA-----AGATTTTATGATGAATACAGCTGAGAGAGGAGCCAGGACTCTGTGT 1768
QY      301 TTTAACAAGC-----AAGAAGCTCAGGAATTTCTGGAACAAATTAATTAATCACTCAAT 356
DB      1769 TTTAATTAAGCCCCCAGAGATCTGTTCACAACAATTAAGCTTTAAAGCTATGCGCCAT 1828
QY      357 CTCTAGATGA--AAATTTTGGAGGAGCTTTGAAGGCTTCAGAAAAGTGTCTCA-- 413
DB      1829 CTGCAAGAGAGAGAAAGTTTGTAAAGGATCTTAAAGGGCTCCAGAGAAATGTTTAA 1888

```

```

QY      414 -ATCAGCTTTAGTACTACAAATAATGCCAGTTTCTTTAAACCAATTTTCACAGT 472
DB      1889 CAGCCTTAGACAAACACTGACAAACATCATGCTTCTTACCAAGCCAAATTTTCACAGT 1948
QY      473 ATCTAAGCTCTTCTGCTCAGAAACCCCGGCTATCATCTTCAAG--GTAAGATT 530
DB      1949 GTTTAATTTCTCCGACTCTCCAAACTCAGCTGCTTTCTTCTTCAAGAGGAGAAAT 2008
QY      531 CAGTGTAGAAATTTCTTCTGAGAACTATTTATTGT 567
DB      2009 CTGTGTAGAAATCTTTATTAAGACATCTCCCTGT 2045

RESULT 11
AB091790
LOCUS      AB091790      6264 bp      DNA      linear      MAM 02-MAY-2003
DEFINITION Bos taurus AMELY gene for amelogenin, partial cds.
ACCESSION  AB091790
VERSION    AB091790.1  GI:29126032
KEYWORDS
SOURCE
ORGANISM   Bos taurus (cow)
            Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE
AUTHORS    1 Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.
TITLE      From the cover: The amelogenin loci span an ancient pseudautosomal
            boundary in diverse mammalian species
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)
MEDLINE    22608569
PUBMED     12672962
REFERENCE
TITLE      2 (bases 1 to 6264)
AUTHORS    Iwase, M., Satta, Y. and Takahata, N.
JOURNAL    Direct Submission
            Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for
            Advanced Studies (Sokendai), Department of Biosystems Science;
            Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan
            (E-mail: iwase@n.koryu.w01.soken.ac.jp, Tel: 81-468-58-1571,
            Fax: 81-468-58-1544)
FEATURES
source      Location/Qualifiers
            1..6264
            /organism="Bos taurus"
            /mol_type="genomic DNA"
            /db_xref="taxon:9913"
            /sex="male"
            114..6264
            /gene="AMELY"
            join(114..169,1585..1596)
            /gene="AMELY"
            join(1597..1659,3515..3562,5080..5124,5863..56264)
            /gene="AMELY"
            /codon_start=1
            /product="amelogenin"
            /protein_id="BAC66109.1"
            /db_xref="GI:29126033"
            /translation="MGTWILFACLVGTAIAMP.PHPRGHAYINFSYVLTPLKX
            YONMLRYPVPSYGVPEVGMHLHQLIPVVSQOPVHIALQPHNHIPVPAQOPVVPQO
            PMPVPGHSMPIOHOPNLPPLPAQOPFOPQOPHOPLOPQPPMHPIOPLPQAPLPPIPLAASGPHAS"
ORIGIN
Query Match      19.1%; Score 129.8; DB 4; Length 6264;
Best local Similarity 62.7%; Pred. No. 3e-18;
Matches 353; Conservative 0; Mismatches 167; Indels 43; Gaps 8;
QY      1 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGACCTTCAGTATCGCTGAGT 60
DB      1597 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGAGCCTATAGTATGCTGTGAGT 1656
QY      61 AAAATTTCCATTTCCATTTTCACAGCTTGGAAATTAATCTGCTCAGATTTTATT 120

```


TITLE PCR detection of the human amelogenin gene and its application to the diagnosis of amelogenesis imperfecta

JOURNAL Bull. Tokyo Dent. Coll. 39 (4), 275-285 (1998)

MEDLINE 99234629

PUBMED 10218009

REFERENCE 2 (bases 1 to 1935)

AUTHORS Sekiguchi, H.

TITLE Direct Submision

JOURNAL Submitted (29-FEB-1996) Hiroshi Sekiguchi, Tokyo Dental College, Pediatric Dentistry, 1-2-2 Masago, Mihama-ku, Chiba, Chiba 261, Japan (Tel:043-270-3945, Fax:043-279-2052)

FEATURES

source

1. .1935
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
1. .1935
/gene="AMGX"
1. .1935
/gene="AMGX"
/note="amelogenin
this nucleotide sequence shows an intron between exon2 and exon3 within human amelogenin gene on the X chromosome"
/number=2

ORIGIN

Query Match 17.4%; Score 118.4; DB 9; Length 1935;
Best Local Similarity 63.7%; Pred. No. 1.1e-15;
Matches 325; Conservative 0; Mismatches 146; Indels 39; Gaps 8;

69 CAATTCCAAATTCACAGCTGGAAATTAATTCGCCACAGTTTATTAGGTTA 128
27 CAGTGTCAAATTCACAACTGACATTAATAATTCGCC-----CATAGTTGCT 75
129 AATTAAAGGTTTAAAGACATCAATCTGATGTTCAAAATGATGTTTAAAGA 188
76 GAAATTAAGGTTTAAAGACATCAATCTGATGTTCAAAATGATGTTTAAAGA 135
189 ACATTTTAAATCTTGATCACAAGATTTTAAACAAAACATTCAGTTGCTTACC 248
136 ACACCTCAGGAGCTTG-----TTTAAAGATATTCCTCAATGCGCGTACC 184
249 CATATCTGATTCAGTATGACAGATGGGGGGTGTGCTGAGAACTCTGAATTTTAA 308
185 AAAA-----ATTCTGATTTGTTACAGCTGGGGGGGCGCCAGACCTCTGATTTTAA 239
309 GCA-----AGAACCTCAGAAATTTCTGGAACAATTAATTAATCATCTCTAGAT 364
240 GCACCCCAGAGATCTGTTGGAACGTGTGCTGTTGAATATCAACACCATCTTAGAT 299
365 GGA- -AAATTTTGAAGGAGCTTTGAAGGCTTCAGAAAAGTGTCAATCAGTCTT 422
300 GGAGGAAGCTTTTGAAGGAGGACCTTGAAG- -GTCTCAGAGAAAGTGTAAACAGCTTT 358
423 ----TAAGTACTACAAAATGCGGCTTTCTTAAACCAATTTCTCAACAAGATTAAC 479
359 GGAACAATATTAAGAGATGCGAGTTTGTCTTAAACCAATTCCTCAAGATTTCAAA 418
480 TCTCTTTCTGCTCAGAAAACCCCGGCTATCATCTTCTCA- -GAGTAAAGATTCAGTGT 537
419 TCTCTTCTGCGCTCCACATATGCTGCTTACCCCTCAGGGGGTAAAGATTTTGTGT 478
538 AGGAATTCCTTCTCTGAATTAATTTGT 567
479 AGGAATTCACCTTTTGAAGCAATTCCTGT 508

RESULT 14
AB091784 7454 bp DNA linear PRI 02-MAY-2003
LOCUS AB091784
DEFINITION Saimiri sciureus AMELY gene for amelogenin, partial cds.
ACCESSION AB091784

VERSION AB091784.1 GI:29126021

KEYWORDS Saimiri sciureus (common squirrel monkey)

SOURCE Saimiri sciureus

ORGANISM Saimiri sciureus

REFERENCE

AUTHORS Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H. and Takahata, N.

TITLE From the Cover: The amelogenin loci span an ancient pseudautosomal boundary in diverse mammalian species

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (9), 5258-5263 (2003)

MEDLINE 22608569

PUBMED 12672962

REFERENCE 2 (bases 1 to 7454)

AUTHORS Iwase, M., Satta, Y. and Takahata, N.

TITLE Direct Submision

JOURNAL Submitted (19-SEP-2002) Mineyo Iwase, Graduate University for Advanced Studies (Sokendai), Department of Biosystems Science; Shonan Kokusaiinura, Hayama, Kanagawa 240-0193, Japan (E-mail: iwase.mineyo@koryu.wol.boken.ac.jp, Tel: 81-468-58-1571, Fax: 81-468-58-1544)

FEATURES

source

1. .7454
/organism="Saimiri sciureus"
/mol_type="genomic DNA"
/db_xref="taxon:9521"
/sex="male"
2028. .7454
/gene="AMELY"
join(2028, 2083, 3075, .3086)
/gene="AMELY"
join(3087, .3140, 5297, .5344, 6741, .6785, 7056, .>7454)
/gene="AMELY"
/codon_start=1
/product="amelogenin"
/protein_id="BAC66104.1"
/db_xref="GI:29126022"
/translation="MGTWLIFACLVVAFAFMPPLPHPGHYINSYEVLTPLKMYOT
MTPSPSYSGYBPMGMLYHQIIPVLSQHPPTHTLQPHHNPVPAQOPVPOOPPM
PVGQHSMPAQHOPNPFPVQOPFOPTVQPOSHQPPQPPVHPVPLPQPPLP
PMPQPLPMPMLPDLMLKGGQ"

ORIGIN

Query Match 14.0%; Score 95.2; DB 9; Length 7454;
Best Local Similarity 58.7%; Pred. No. 1e-10;
Matches 223; Conservative 0; Mismatches 148; Indels 9; Gaps 3;

197 AAAATCTTATCAAGATTTTAAACAAAACATTCAGTTGCTTACCCATATTCT 256
3505 AAAAATTAATAATCAATTAATAATAATAATGCTTTTATGATGCTTCACAGA 3564
257 GATTCAGTATGACAGAGTGGGGGTGTGCTGAGAACTGGAATTTTAAACAAGACA 316
3565 GGTTCGATTTGTTAGTACGTGAGGTTGGGCCATATCTGAATTTTAAAGACCCG 3624
317 CTCAGAAATCTTGAACAATTAATTAAT- -CACTCATCTTAGATGAAAAT 372
3625 GAGTTCTGTGACACCATTAAGCTGTATTAATTAACAACCATTTGATGAGAAAGCT 3684
373 TTTGAAGGAGCACTTGAAGAGCTCCAGAAAAGTGTCAATTCAGTCTT- -TAAGTAC 429
3685 TTTGAAGGATCTTTGAAGATCTTCACAGAAAGCTTAATTTAGCCTTAATAATAAT 3744
430 TACAAAATGCGAGTTTCTTAAACCAATTTCTCAAGATTAATCTCTTTCTG 489
3745 AACAAAATACACAGTTTGTCTTAAACGCAATTCCTCAATTTCCAAATCTTTCTTG 3804
490 CTCAGAAAACCCCGGCTATCATCTTTC- -AGAGTAAGATTCAGTGTAGAAATTTCC 547
3805 CCCTCCACATATTAATAGCTTATCTCTCAGAGAGTAAGATTTCTGTGTAGAAACCCAC 3864
548 TTCCTGAATTAATTTGT 567

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 11:25:10 ; Search time 902.049 Seconds
(without alignments)
3957.220 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680
1 atggggagactgagcttcttgc.....ggtactgagagagagacac 680

Sequence: 1 atggggagactgagcttcttgc.....ggtactgagagagagacac 680

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160.4	23.6	259202	12	ADQ18492 Human sof
2	51.2	7.5	722	6	AA141111
3	51.2	7.5	752	6	AA141110
4	49	7.2	270	12	ADM80835
5	49	7.2	349	12	ADM80826
6	49	7.2	623	10	ADM80827
7	45.6	6.7	556	10	ADM80826
8	45.6	6.7	556	10	ADM80827
9	45	6.6	476	3	AAZ50832
10	44.8	6.6	318	12	ADM80836
11	44.8	6.6	750	2	AAZ07020
12	44.8	6.6	793	2	AAZ07018
13	43.2	6.4	802	2	AAZ07019
14	43.2	6.4	852	12	ADQ22977
15	43	6.3	178896	6	ABO88146
16	41.2	6.1	2000	6	ABZ17518
17	41.2	6.1	15399	6	ABL33515
18	40.6	6.0	2000	8	ADA71938
19	40.4	5.9	110000	5	AA161373
20	39.8	5.9	3001	6	ABK34028
21	39.8	5.9	5118	8	ADA20376

22	39.8	5.9	5118	8	ADA84183	ADA84183 Human ren
23	39.8	5.9	7763	6	ABL34436	ABL34436 Human imm
24	39.8	5.9	110000	6	ABA03041_00	ABA03041 Listeria
25	39.6	5.8	645	4	AA108580	AA108580 Human bre
26	39.6	5.8	3474	4	ABL14636	ABL14636 Drosophila
27	39.6	5.8	7113	6	ABL132805	ABL132805 Human imm
28	39.6	5.8	8979	6	ABL32785	ABL32785 Human imm
29	39.6	5.8	8979	6	ABK31271	ABK31271 Signal tr
30	39.6	5.8	8979	6	ABL70232	ABL70232 Chemical
31	39.6	5.8	8979	6	AA561179	AA561179 Human gen
32	39.6	5.8	80331	4	AA089559	AA089559 Human his
33	39.2	5.8	6282	6	ABL32455	ABL32455 Human imm
34	39.2	5.8	6641	6	ABN80003	ABN80003 Human che
35	39	5.7	65	6	ABN28389	ABN28389 Rat eplic
36	39	5.7	2625	4	ABL24966	ABL24966 Drosophila
37	39	5.7	2780	12	AD143260	AD143260 Plant tra
38	39	5.7	5964	6	ABN80170	ABN80170 Human che
39	39	5.7	8456	6	ABL33976	ABL33976 Human imm
40	38.6	5.7	2478	2	AA095552	AA095552 Nucleic a
41	38.6	5.7	8056	8	AB210246	AB210246 Haematopo
42	38.6	5.7	110000	12	AD079173_0	AD079173 KLF12 gen
43	38.4	5.6	492	10	ADB81854	ADB81854 Arabidops
44	38.4	5.6	11622	6	ABL32674	ABL32674 Human imm
45	38.4	5.6	16228	6	ABL70459	ABL70459 Chemical

ALIGNMENTS

RESULT 1	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
ID	ADQ18492	ADQ18492 standard; DNA; 259202 BP.
XX	ADQ18492:	
AC	ADQ18492:	
XX	26-AUG-2004 (first entry)	
DT	26-AUG-2004 (first entry)	
XX	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
DE	Human soft tissue sarcoma-upregulated DNA - SEQ ID 1311.	
XX	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human; de.	
KW	de.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	WO2004048938-A2.	
PN	WO2004048938-A2.	
XX	10-JUN-2004.	
PD	10-JUN-2004.	
XX	26-NOV-2003; 2003WO-US038193.	
PF	26-NOV-2003; 2003WO-US038193.	
XX	26-NOV-2002; 2002US-0429739P.	
PR	26-NOV-2002; 2002US-0429739P.	
XX	(PROT-) PROTEIN DESIGN LABS INC.	
PA	(PROT-) PROTEIN DESIGN LABS INC.	
XX	Aziz N, Ginsburg WM, Zlotnick A;	
PI	Aziz N, Ginsburg WM, Zlotnick A;	
XX	WPI; 2004-441208/41.	
DR	WPI; 2004-441208/41.	
XX	Early detection of soft tissue sarcoma comprises determining expression	
PT	Early detection of soft tissue sarcoma comprises determining expression	
XX	of a gene in a first soft tissue sample and a normal soft tissue sample	
PT	of a gene in a first soft tissue sample and a normal soft tissue sample	
XX	and comparing the gene expression, also useful in treating soft tissue	
PT	and comparing the gene expression, also useful in treating soft tissue	
XX	sarcoma.	
PT	sarcoma.	
XX	Example 2; SEQ ID NO 1311; 210pp; English.	
PS	Example 2; SEQ ID NO 1311; 210pp; English.	
XX	The invention relates to a novel method for detecting soft tissue sarcoma	
XX	which comprises obtaining a first soft tissue sample from an individual	
CC	and a normal soft tissue sample from the same or different individual,	
CC	determining the expression of a gene in both samples and comparing the	
CC	expression of the gene in both soft tissue samples, where a higher level	
CC	of protein expression in the first soft tissue sample indicates the	
CC	presence of soft tissue sarcoma. The method of the invention has	
CC	cytostatic applications and may be useful for detecting soft tissue	

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma up-regulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX Sequence 259202 BP; 81699 A; 51421 C; 49221 G; 76861 T; 0 U; 0 Other;

Query Match 23.6%; Score 160.4; DB 12; Length 259202;
 Best Local Similarity 65.1%; Pred. No. 3.3e-31;
 Matches 384; Conservative 0; Mismatches 156; Indels 50; Gaps 8;

QY 1 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGAGCTTCAGTATGCTGTAGT 60
 Db ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGAGCTTCAGTATGCTGTAGT 60932
 QY 61 AAAATTC-----CAATTCATTTTACAGCTTGGAAATTAATTCGCT 108
 Db AAAACACCCCTTGATAGTCAAGTCCAAATTCACAACTTGGACATTAATAATCTGC-- 69990
 QY 109 CAGATTTTATTTAGGGTTAAATTTAGGGTTTAAAGCAGTACAGATCTGATGTCACA 168
 Db 69991 -----TCATAGTTGCTGAATTAAGGTTTAAACAGTACAGATCTGCTTCA 70041
 QY 169 AATGTGACTGTGTTTAAAGAAACATTTTAAATCTGATCACAAGATTTTAAACAAAA 228
 Db 70042 TATGCTCTGGGTTGAAGAAACACTTCAGAGACT-----CTTTTAAAGGTA 70090
 QY 229 CATTTCTAGTTGCTTCAACCATATTTGATTCAGTATGACAGAGTGGGGGCTGTGCTG 288
 Db 70091 TATTTCTAAATGCGCTACCAAAA-----ATTCTGATTTGGTACAGCTGGGGGCGGCCCC 70145
 QY 289 AGAAGCTGAAATTTTAAACAAGCA---AGAACCTCAGAAATCTTGAAACAATTACTT 344
 Db 70146 AGAGCTCTGATTTTAAAGACCCCAAGAGATCTGTTGGAAGCTTATGCTTGAAT 70205
 QY 345 ATAACTACTATCTCTAGATGA--AAATTTTGAAGGAGCCTTGAAGGCTCCAGA 402
 Db 70206 ATCACACCCCTCTCTAGATGAGAGAGCTTTGGAAGGAGCCTTGAAGGCTCCAGA 70265
 QY 403 AAAAGTCTAATGAGCTT---AAGTACTACAAAATGCCGTTTCTCTTAAACCA 459
 Db 70266 GAAAGTCTTAACAGCTTGGACAAATTAACAGATGCCAGTTTGTCTTAAACCA 70325
 QY 460 ATTTCACAAGTATCTACTCTCTCTGCTCAGAAAACCCCGGCTATCATTTCTTCA 519
 Db 70326 ATTCTCTCAAGATTCAAATCTCTCTGCTCCCAATATGCTCTTACCCCTCA 70385
 QY 520 --GAGTAAAGATTCAGTGTAGGAATTCCTTCTGAACTATTTATTTGT 567
 Db 70386 GGGGGTAAAGATTTTGTGTAGGATTCACCTTTTGAAGCACAATTCCTGT 70435

RESULT 2

AA141111 standard; DNA; 722 BP.

AA141111;

16-OCT-2002 (first entry)

gAML related Y-chromosome DNA sequence.

Goat embryo sexual identification technique; goat amelogenin gene; gAML;
 sex-specific; gene; ds; Y-chromosome.

Capra hircus.

Key Location/Qualifiers
 CDS 35..658
 /tag= a
 /product= "Y-chromosome protein"

PN TW454013-A.

PD 11-SEP-2001.

PF 10-NOV-1999; 99TW-00119616.

PR 10-NOV-1999; 99TW-00119616.

XX (CHEN/) CHEN C.
 XX (JANG/) JANG J.
 XX (WENG/) WENG T.
 XX (JENG/) JENG D.

PI Chen C, Jang J, Weng T, Jeng D;

DR WPI; 2002-442016/47.

DR P-PSDB; AAO22534.

PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
 PT identification, comprises high sensitivity even using single white blood
 PT cell or cleavage c.

PS Disclosure; Page 28; 35pp; Chinese.

XX The invention relates to a goat embryo sexual identification technique
 CC with high efficiency, sensitivity and repeatability. This technique
 CC involves separately cloning and sequencing the coding regions and the
 CC introns of the goat amelogenin gene (gAML) on the goat chromosomes. The
 CC results indicate that there are sex-specific sequences in the fifth
 CC intron of the gene. The major characteristics according to the present
 CC invention include high sensitivity, applicable in sex identification even
 CC only using a single white blood cell or a single cleavage cell of
 CC blastula; high diagnostic efficiency, capable of identifying hundreds of
 CC goat embryo in 3 hours; simple operation procedures without complicated
 CC steps of DNA extraction and need no additional control group intron; and
 CC can be applied on different species of goats. This polynucleotide
 CC sequence represents a gAML related Y-chromosome DNA sequence of the
 CC invention

SQ Sequence 722 BP; 173 A; 264 C; 140 G; 145 T; 0 U; 0 Other;

Query Match 7.5%; Score 51.2; DB 6; Length 722;

Best Local Similarity 94.6%; Pred. No. 0.0025;

Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGAGCTTCAGTATGCTGT 56
 Db 35 ATGGGACCTGGATTTGTTGCTGCTGCTGAGACAGAGCTTCAGTATGCTGT 90

RESULT 3

AA141110 standard; DNA; 752 BP.

AA141110;

16-OCT-2002 (first entry)

gAML related X-chromosome DNA sequence.

Goat embryo sexual identification technique; goat amelogenin gene; gAML;
 sex-specific; gene; ds; X-chromosome.

Capra hircus.

Key Location/Qualifiers
 CDS 35..658
 /tag= a
 /product= "X-chromosome protein"

TW454013-A.

PD 11-SEP-2001.

```
XX PF 10-NOV-1999; 99TW-00119616.
XX PR 10-NOV-1999; 99TW-00119616.
XX PA (CHEN/) CHEN C.
XX PA (JANG/) JANG J.
XX PA (WENG/) WENG T.
XX PA (JENG/) JENG D.
XX PI Chen C, Jang J, Weng T, Jeng D;
XX WPI: 2002-442016/47.
XX DR P-PSDB; AAO22534.
XX PT Sex-specific sequence of goat amelogenin gene, useful for embryo sexual
XX PT identification, comprises high sensitivity even using single white blood
XX PT cell or cleavage c.
XX PS Disclosure; Page 28; 35pp; Chinese.
XX CC The invention relates to a goat embryo sexual identification technique
XX CC with high efficiency, sensitivity and repeatability. This technique
XX CC involves separately cloning and sequencing the coding regions and the
XX CC introns of the goat amelogenin gene (GAML) on the goat chromosomes. The
XX CC results indicate that there are sex-specific sequences in the fifth
XX CC intron of the gene. The major characteristics according to the present
XX CC invention include high sensitivity, applicability in sex identification even
XX CC only using a single white blood cell or a single cleavage cell of
XX CC blastula, high diagnostic efficiency, capable of identifying hundreds of
XX CC goat embryo in 3 hours; simple operation procedures without complicated
XX CC steps of DNA extraction and need no additional control group intron; and
XX CC can be applied on different species of goats. This polynucleotide
XX CC sequence represents a GAML related X-chromosome DNA sequence of the
XX CC invention
XX SQ Sequence 752 BP; 184 A; 271 C; 152 G; 145 T; 0 U; 0 Other;
XX
XX Query Match 7.5%; Score 51.2; DB 6; Length 752;
XX Best Local Similarity 94.6%; Pred. No. 0.0026;
XX Matches 53; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGGACCTGATTTGTTGCTGCTGCTGGAGACAGCTTCAATATGCCGT 56
DB 35 ATGGGGACCTGATTTGTTGCTGCTGCTGGAGACAGCTTCTATGCCCT 90
XX
XX RESULT 4
XX ADM80835
XX ID ADM80835 standard; cDNA; 270 BP.
XX AC ADM80835;
XX DT 03-JUN-2004 (first entry)
XX DE Human CADECM-22 encoding cDNA SEQ ID NO:64.
XX XX human, cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangler disease; gene; ss.
XX OS Homo sapiens.
XX XX
XX Key Location/Qualifiers
XX FH CDS 36..212
XX FT /*tag= a
XX FT /product= "CADECM-22"
XX PN WO2004015396-A2.
XX PD 19-FEB-2004.
```

```
XX PF 12-AUG-2003; 2003WO-US025418.
XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424504P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX PA (INCY-) INCYTE CORP.
XX XX
XX PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
XX PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;
XX WPI: 2004-191795/18.
XX DR P-PSDB; ADM80793.
XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.
XX PS Claim 5; SEQ ID NO 64; 272pp; English.
XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADECM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative
XX CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX CC cancer, obesity and Tangler disease.
XX SQ Sequence 270 BP; 76 A; 65 C; 67 G; 62 T; 0 U; 0 Other;
XX
XX Query Match 7.2%; Score 49; DB 12; Length 270;
XX Best Local Similarity 79.5%; Pred. No. 0.0071;
XX Matches 58; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1 ATGGGGACCTGATTTGTTGCTGCTGCTGGAGACAGCTTCAATATGCCGTGAGT 60
DB 36 ATGGGGACCTGATTTATTTATTTGCTGCTGCTGCTGGAGACAGCTTTGCCATGCTGCTT 95
QY 61 AAAATTTCCAATT 73
DB 96 ACCCCTTGAAGT 108
XX
XX RESULT 5
XX ADM80826
XX ID ADM80826 standard; cDNA; 549 BP.
XX AC ADM80826;
XX DT 03-JUN-2004 (first entry)
XX DE Human CADECM-13 encoding cDNA SEQ ID NO:55.
XX XX human, cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangler disease; gene; ss.
XX OS Homo sapiens.
XX XX
XX Key Location/Qualifiers
XX FH CDS 36..491
XX FT /*tag= a
XX FT /product= "CADECM-13"
XX PD 19-FEB-2004.
```

PN WO2004015396-A2.
 XX
 PD 19-FEB-2004.
 XX
 PF 12-AUG-2003; 2003WO-US025418.
 XX
 PR 13-AUG-2002; 2002US-0403781P.
 PR 30-AUG-2002; 2002US-0407034P.
 PR 13-SEP-2002; 2002US-0410566P.
 PR 24-SEP-2002; 2002US-0413482P.
 PR 25-SEP-2002; 2002US-0413890P.
 PR 08-NOV-2002; 2002US-0424984P.
 PR 13-NOV-2002; 2002US-0426222P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
 PI Wang JT, Chien D, Yang YG;
 XX
 DR WPI; 2004-191795/18.
 DR P-PSDB; ADM80784.
 XX
 PT New cell adhesion and extracellular matrix proteins, useful in
 PT diagnosing, treating and preventing immune, neurological, developmental,
 PT connective tissue and cell proliferative disorders including cancer.
 XX
 PS Claim 5; SEQ ID NO 55; 272pp; English.
 XX
 CC The present sequence encodes a human cell adhesion and extracellular
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
 CC cytoskeletal and anorectic activities. The CADECM polypeptides and
 CC polynucleotides are useful in diagnosing, treating and preventing immune,
 CC neurological, developmental, connective tissue and cell proliferative
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
 CC cancer, obesity and Tangier disease.
 XX
 SQ Sequence 549 BP, 134 A; 203 C; 112 G; 100 T; 0 U; 0 Other;
 XX

Query Match 7.2%; Score 49; DB 12; Length 549;
 Best Local Similarity 79.5%; Pred. No. 0.0088;
 Matches 59; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGGAGACGCTTCAGTATGCTGTAGT 60
 DB 36 ATGGGACCTGATTTGTTGCTGCTGCTGGAGACGCTTCAGTATGCTGTAGT 95
 QY 61 AAAATTTCCAATT 73
 DB 96 ACCCCTTTGAAAGT 108

RESULT 6
 ADM80827
 ID ADM80827 standard; cDNA; 623 BP.
 XX
 AC ADM80827;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human CADECM-14 encoding cDNA SEQ ID NO:56.
 XX
 XX human; cell adhesion and extracellular matrix protein; CADECM;
 XX neuroprotective; cytoskeletal; anorectic; immune disorder;
 XX neurological disorder; developmental disorder;
 XX connective tissue disorder; cell proliferative disorder; cancer; obesity;
 XX Tangier disease; gene; ss.
 XX
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX FT CDS 36..563

PT /*tag= a
 FT /product= "CADECM-14"
 XX
 PN WO2004015396-A2.
 XX
 PD 19-FEB-2004.
 XX
 PF 12-AUG-2003; 2003WO-US025418.
 XX
 PR 13-AUG-2002; 2002US-0403781P.
 PR 30-AUG-2002; 2002US-0407034P.
 PR 13-SEP-2002; 2002US-0410566P.
 PR 24-SEP-2002; 2002US-0413482P.
 PR 25-SEP-2002; 2002US-0413890P.
 PR 08-NOV-2002; 2002US-0424984P.
 PR 13-NOV-2002; 2002US-0426222P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
 PI Becha SD, Margulis JP, Swarnakar A, Chawla NK, Ramkumar J;
 PI Hafalia AJA, Lee SY, Jiang X, Jackson AA, Richardson TW, Blake JJ;
 PI Wang JT, Chien D, Yang YG;
 XX
 DR WPI; 2004-191795/18.
 DR P-PSDB; ADM80785.
 XX
 PT New cell adhesion and extracellular matrix proteins, useful in
 PT diagnosing, treating and preventing immune, neurological, developmental,
 PT connective tissue and cell proliferative disorders including cancer.
 XX
 PS Claim 5; SEQ ID NO 56; 272pp; English.
 XX
 CC The present sequence encodes a human cell adhesion and extracellular
 CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
 CC cytoskeletal and anorectic activities. The CADECM polypeptides and
 CC polynucleotides are useful in diagnosing, treating and preventing immune,
 CC neurological, developmental, connective tissue and cell proliferative
 CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
 CC cancer, obesity and Tangier disease.
 XX
 SQ Sequence 623 BP; 150 A; 228 C; 128 G; 117 T; 0 U; 0 Other;
 XX

Query Match 7.2%; Score 49; DB 12; Length 623;
 Best Local Similarity 79.5%; Pred. No. 0.0092;
 Matches 58; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTGTTGCTGCTGCTGGAGACGCTTCAGTATGCTGTAGT 60
 DB 36 ATGGGACCTGATTTGTTGCTGCTGCTGGAGACGCTTCAGTATGCTGTAGT 95
 QY 61 AAAATTTCCAATT 73
 DB 96 ACCCCTTTGAAAGT 108

RESULT 7
 ADB59026
 ID ADB59026 standard; DNA; 556 BP.
 XX
 AC ADB59026;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Toxicity-related gene, SEQ ID 4052.
 XX
 XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
 XX drug screening; toxicity assay; ds.
 XX
 XX Unidentified.
 XX
 XX WO2003064624-A2.

PD 07-AUG-2003.
XX
PF 31-JAN-2003; 2003MO-US003194.
XX
PR 31-JAN-2002; 2002US-00060087.
XX
PR 15-MAR-2002; 2002US-0364045P.
XX
PR 15-MAR-2002; 2002US-0364055P.
XX
PR 30-DEC-2002; 2002US-0436643P.
XX
(GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Caetle A, Elashoff M;
XX WPI; 2003-689530/65.
XX
DR
XX
PT Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
PT the compound.
XX
PS Claim 1; SEQ ID NO 4052; 1156bp; English.
XX
CC The present invention relates to a method for predicting a toxic effect
CC of a compound. The method comprises preparing a gene expression profile
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in
CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
XX
Query Match 6.7%; Score 45.6; DB 10; Length 556;
Best Local Similarity 90.6%; Pred. No. 0.069;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGCAGCTTGTATGCC 53
DB 22 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGCAGCTTGTATGCC 74
XX
RESULT 8
ID ADB53782 standard; DNA; 556 BP.
XX
AC ADB53782;
XX
DT 04-DEC-2003 (first entry)
XX
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4324.
XX
KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KM toxicity marker; toxicity progression; drug screening;
KM primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS Rattus norvegicus.
XX
PN MO2003065993-A2.
XX
PD 14-AUG-2003.
XX
PF 04-FEB-2003; 2003MO-US003482.
XX

PR 04-FEB-2002; 2002US-0351171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
(GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Caetle A, Orr M;
PI Elashoff M;
XX
DR WPI; 2003-731472/69.
XX
XX
PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
PS Claim 44; SEQ ID NO 4324; 874bp; English.
XX
XX
CC The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
SQ Sequence 556 BP; 121 A; 183 C; 114 G; 119 T; 0 U; 19 Other;
XX
Query Match 6.7%; Score 45.6; DB 10; Length 556;
Best Local Similarity 90.6%; Pred. No. 0.069;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGCAGCTTGTATGCC 53
DB 22 ATGGGAGCTGATTTTGTTCCTGCTCTCGGAGCAGCTTGTATGCC 74
XX
RESULT 9
ID AAZ50832 standard; DNA; 476 BP.
XX
AC AAZ50832;
XX
DT 31-MAY-2000 (first entry)
XX
DE Rat amelogenin gene (A4).
XX
XX
KM Amelogenin; splice variant; rat; (A4); chondrogenesis; osteogenesis;
KM chondrogenic inducing molecule; CIM; cartilage growth; osteopathic;
KM extracellular matrix protein; tooth enamel; enamel mineralisation;
KM ameloblast; bone regeneration; composite cell construct; ds.
XX

```

XX OS Rattus sp.
XX
XX Key Location/Qualifiers
XX FT exon 1..36
XX FT exon /*tag= a
XX FT exon /partial
XX FT exon /number= 1
XX FT exon 37..101
XX FT exon /*tag= b
XX FT exon /number= 2
XX FT exon 48..317
XX FT exon /*tag= c
XX FT exon /product= "Rat amelogenin protein"
XX FT exon /transl_except= (pos:69..71, aa:Gly)
XX FT exon 48..95
XX FT exon /*tag= d
XX FT exon 96..314
XX FT exon /*tag= e
XX FT exon /label= Mature_rat_amelogenin_protein
XX FT exon 102..149
XX FT exon /*tag= f
XX FT exon /number= 3
XX FT exon 150..191
XX FT exon /*tag= g
XX FT exon /number= 4
XX FT exon 192..236
XX FT exon /*tag= h
XX FT exon /number= 5
XX FT exon 237..311
XX FT exon /*tag= i
XX FT exon /number= 6d
XX FT exon /note= "Comprises of gene segments 6a, b, c and d"
XX FT exon 312..317
XX FT exon /*tag= j
XX FT exon /number= 7
XX FT exon /note= "Includes the stop codon"
XX PN WO200006734-A1.
XX
XX PD 10-FEB-2000.
XX
XX PF 29-JUL-1999; 99WO-US017242.
XX
XX PR 29-JUL-1998; 98US-0054489P.
XX
XX PA (NOUN ) UNIV NORTHWESTERN.
XX
XX PI Vels A, Nebgen DR;
XX
XX DR WPI; 2000-205464/18.
XX
XX DR P-PSDB; AAY45074.
XX
XX PT Novel amelogenin polypeptides and polynucleotides, useful for enhancing
XX PT bone generation in mammals and synthesizing bone matrix or articular
XX PT surfaces at implant sites.
XX
XX PS Example 2; Fig 11B; 79pp; English.
XX
XX CC The present DNA sequence is the full-length rat amelogenin gene (A4),
XX CC comprising exons 1-7, including the exon segment 6d. It is derived from
XX CC the rat incisor odontoblast-pulp cDNA library. The splice variants of
XX CC this gene functions as an osteogenic or chondrogenic inducing molecule
XX CC (CIM), which is useful for enhancing bone or cartilage growth. It has
XX CC osteoparathic activity. Amelogenin belongs to the family of extracellular
XX CC matrix proteins, in developing tooth enamel, that are produced by the
XX CC ameloblasts and plays a role in enamel mineralisation. Chondrogenic or
XX CC osteogenic inducing amelogenin molecules are useful to induce
XX CC differentiation of cells to the osteogenic and chondrogenic phenotypes
XX CC and can be used in a composite cell construct for bone and cartilage
XX CC regeneration. The polynucleotides can be employed to produce the
XX CC polypeptides by recombinant techniques

```

```

SQ Sequence 476 BP; 155 A; 106 C; 98 G; 117 T; 0 U; 0 Other;
Query Match 6.6%; Score 45; DB 3; Length 476;
Best Local Similarity 90.6%; Pred. No. 0.095;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY
1 ATGGGACCTGGATTGTTTTCCTGCTCTGAGGAGCAGCCTTGTATGTC 53
48 ATGGGACCTGGATTGTTTTCCTGCTCTGAGGAGCAGCCTTGTATGTC 100
Db
RESULT 10
ADM80836
ID ADM80836 standard; cDNA; 318 BP.
XX
XX AC ADM80836;
XX
XX DT 03-JUN-2004 (first entry)
XX
XX DE Human CADECM-23 encoding cDNA SEQ ID NO:65.
XX
XX KW human; cell adhesion and extracellular matrix protein; CADECM;
XX KW neuroprotective; cytoskeletal; anorectic; immune disorder;
XX KW neurological disorder; developmental disorder;
XX KW connective tissue disorder; cell proliferative disorder; cancer; obesity;
XX KW Tangier disease; gene; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 36..260
XX FT /*tag= a
XX FT /product= "CADECM-23"
XX
XX PN WO2004015396-A2.
XX
XX PD 19-FEB-2004.
XX
XX PF 12-AUG-2003; 2003WO-US025418.
XX
XX PR 13-AUG-2002; 2002US-0403781P.
XX PR 30-AUG-2002; 2002US-0407034P.
XX PR 13-SEP-2002; 2002US-0410566P.
XX PR 24-SEP-2002; 2002US-0413482P.
XX PR 25-SEP-2002; 2002US-0413890P.
XX PR 08-NOV-2002; 2002US-0424804P.
XX PR 13-NOV-2002; 2002US-0426222P.
XX
XX PA (INCY-) INCYTE CORP.
XX
XX PI Elliott VS, Khare R, Emerling BM, Kable AE, Tran UK, Jin P;
XX PI Becha SD, Margis JP, Swarnakar A, Chawla NK, Ramkumar J;
XX PI Hafalia AJA, Lee SY, Ujang X, Jackson AA, Richardson TW, Blake JJ;
XX PI Wang JT, Chien D, Yang YG;
XX
XX DR WPI; 2004-191795/18.
XX
XX DR P-PSDB; ADM80794.
XX
XX PT New cell adhesion and extracellular matrix proteins, useful in
XX PT diagnosing, treating and preventing immune, neurological, developmental,
XX PT connective tissue and cell proliferative disorders including cancer.
XX
XX PS Claim 5; SEQ ID NO 65; 272pp; English.
XX
XX CC The present sequence encodes a human cell adhesion and extracellular
XX CC matrix protein designated CADECM. CADECM sequences has neuroprotective,
XX CC cytoskeletal and anorectic activities. The CADECM polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing immune,
XX CC neurological, developmental, connective tissue and cell proliferative
XX CC disorders including cancer, e.g. breast, prostate, ovarian, lung or colon
XX CC cancer, obesity and Tangier disease.
XX
XX SQ Sequence 318 BP; 87 A; 81 C; 75 G; 75 T; 0 U; 0 Other;

```

Query Match 6.6%; Score 44.8; DB 12; Length 318;
Best Local Similarity 87.5%; Pred. No. 0.095; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTTCCTGCTCGGAGAGACCTTCAGTATGCTGT 56
DB 36 ATGGGACCTGATTTTATTTGCTGCTCGGAGAGACCTTTGCAATGCTCT 91

RESULT 11
AAZ07020
ID AAZ07020 standard; DNA; 750 BP.
XX
XX AAZ07020;
XX
XX 15-NOV-1999 (first entry)
XX
XX Amelogenin X nucleotide sequence.
XX
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
XX multiple allelic site; apolipoprotein E; apof; coronary artery disease;
XX Alzheimer's disease; ds.
XX
XX Unidentified.
XX
XX MO9940226-A2.
XX
XX 12-AUG-1999.
XX
XX 08-JAN-1999; 99MO-US000499.
XX
XX 04-FEB-1998; 98US-00018595.
XX
XX (PEKE) PERKIN-ELMER CORP.
XX
XX Livak KJ, Goodsaiaid F;
XX
XX MPI; 1999-539985/45.
XX
XX 5' nuclease amplification assay using fluorescence-quencher probes for
XX determination of a genotype at multiple allelic sites.
XX
XX Disclosure; Fig 10; 95pp; English.
XX
XX The present invention describes first and second sets of fluorescer-
XX quencher probes used simultaneously in a 5' nuclease assay to identify
XX which members of a first or second set of substantially homologous
XX sequences are present in a DNA sample. The method can be used to genotype
XX a sample of genomic DNA at two or more different allelic sites.
XX Generating a fluorescence spectrum and signature for each genotype, which
XX uniquely reflects the assay's inherent inefficiency for that genotype
XX given the particular conditions, probes and primers used, the genotype of
XX unknown sequences can be determined. The assay was shown to be useful for
XX determining apof genotypes. The assay can be used as a diagnostic tool
XX for assessing the risk for coronary artery disease and/or late-onset
XX Alzheimer's disease. Using the 5' nuclease assay of the invention it is
XX possible to determine a genotype at two or more allelic sites in a single
XX reaction. This approach is much faster than previous approaches to
XX genotyping genes having two or more allelic sites, such as the
XX apolipoprotein E gene. A key advantage of the method for determining the
XX genotype of a sample of DNA at multiple allelic sites is that it does not
XX rely on 5' nuclease assay working with 100% efficiency to distinguish
XX between substantially homologous sequences such as alleles. The present
XX sequence represent the nucleotide sequence for amelogenin X, which is
XX used in the exemplification of the present invention

SO Sequence 750 BP; 191 A; 260 C; 146 G; 153 T; 0 U; 0 Other;

Query Match 6.6%; Score 44.8; DB 2; Length 750;
Best Local Similarity 87.5%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTTCCTGCTCGGAGAGACCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGATTTTATTTGCTGCTCGGAGAGACCTTTGCCAATGCTCT 124

RESULT 12
AAZ07018
ID AAZ07018 standard; DNA; 793 BP.
XX
XX AAZ07018;
XX
XX 15-NOV-1999 (first entry)
XX
XX Amelogenin X nucleotide sequence.
XX
XX Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
XX multiple allelic site; apolipoprotein E; apof; coronary artery disease;
XX Alzheimer's disease; ss.
XX
XX Unidentified.
XX
XX MO9940226-A2.
XX
XX 12-AUG-1999.
XX
XX 08-JAN-1999; 99MO-US000499.
XX
XX 04-FEB-1998; 98US-00018595.
XX
XX (PEKE) PERKIN-ELMER CORP.
XX
XX Livak KJ, Goodsaiaid F;
XX
XX MPI; 1999-539985/45.
XX
XX 5' nuclease amplification assay using fluorescence-quencher probes for
XX determination of a genotype at multiple allelic sites.
XX
XX Disclosure; Fig 8A; 95pp; English.
XX
XX The present invention describes first and second sets of fluorescer-
XX quencher probes used simultaneously in a 5' nuclease assay to identify
XX which members of a first or second set of substantially homologous
XX sequences are present in a DNA sample. The method can be used to genotype
XX a sample of genomic DNA at two or more different allelic sites.
XX Generating a fluorescence spectrum and signature for each genotype, which
XX uniquely reflects the assay's inherent inefficiency for that genotype
XX given the particular conditions, probes and primers used, the genotype of
XX unknown sequences can be determined. The assay was shown to be useful for
XX determining apof genotypes. The assay can be used as a diagnostic tool
XX for assessing the risk for coronary artery disease and/or late-onset
XX Alzheimer's disease. Using the 5' nuclease assay of the invention it is
XX possible to determine a genotype at two or more allelic sites in a single
XX reaction. This approach is much faster than previous approaches to
XX genotyping genes having two or more allelic sites, such as the
XX apolipoprotein E gene. A key advantage of the method for determining the
XX genotype of a sample of DNA at multiple allelic sites is that it does not
XX rely on 5' nuclease assay working with 100% efficiency to distinguish
XX between substantially homologous sequences such as alleles. The present
XX sequence represent the nucleotide sequence for amelogenin X, which is
XX used in the exemplification of the present invention

SO Sequence 793 BP; 215 A; 266 C; 150 G; 162 T; 0 U; 0 Other;

Query Match 6.6%; Score 44.8; DB 2; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTTCCTGCTCGGAGAGACCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGATTTTATTTGCTGCTCGGAGAGACCTTTGCCAATGCTCT 124

```
RESULT 13
AAZ07019
ID AAZ07019 standard; DNA; 802 BP.
XX
AC AAZ07019;
XX
DT 15-NOV-1999 (first entry)
XX
DE Amelogenin Y nucleotide sequence.
XX
KM Amelogenin X; Amelogenin Y; genotype; 5' nuclease amplification;
XX multiple allele site; apolipoprotein E; apoE; coronary artery disease;
XX Alzheimer's disease; ss.
XX
OS Unidentified.
XX
XX WO940226-A2.
XX
PD 12-AUG-1999.
XX
PF 08-JAN-1999; 99MO-US000499.
XX
PR 04-FEB-1998; 98US-00018595.
XX
PA (PEKE ) PERKIN-ELMER CORP.
XX
PI Liyak KJ, Goodaaid F;
XX
DR WPI; 1999-539985/45.
XX
PT 5' nuclease amplification assay using fluorescence-quencher probes for
XX determination of a genotype at multiple allelic sites.
XX
PS Disclosure; Fig 8B; 95pp; English.
XX
XX
CC The present invention describes first and second sets of fluorescer-
CC quencher probes used simultaneously in a 5' nuclease assay to identify
CC which members of a first or second set of substantially homologous
CC sequences are present in a DNA sample. The method can be used to genotype
CC a sample of genomic DNA at two or more different allelic sites.
CC
CC Generating a fluorescence spectrum and signature for each genotype, which
CC uniquely reflects the assay's inherent inefficiency for that genotype
CC given the particular conditions, probes and primers used, the genotype of
CC unknown sequences can be determined. The assay was shown to be useful for
CC determining apoe genotypes. The assay can be used as a diagnostic tool
CC for assessing the risk for coronary artery disease and/or late-onset
CC Alzheimer's disease. Using the 5' nuclease assay of the invention it is
CC possible to determine a genotype at two or more allelic sites in a single
CC reaction. This approach is much faster than previous approaches to
CC genotyping genes having two or more allelic sites, such as the
CC apolipoprotein E gene. A key advantage of the method for determining the
CC genotype of a sample of DNA at multiple allelic sites is that it does not
CC rely on 5' nuclease assay working with 100% efficiency to distinguish
CC between substantially homologous sequences such as alleles. The present
CC sequence represent the nucleotide sequence for amelogenin Y, which is
CC used in the exemplification of the present invention.
XX
SQ Sequence 802 BP; 216 A; 258 C; 155 G; 173 T; 0 U; 0 Other;
XX
Query Match 6.4%; Score 43.2; DB 2; Length 802;
Best Local Similarity 85.7%; Pred. No. 0.33;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
QY 1 ATGGGACCTGGATTGTTGCTGCTGCGAGACGACCTTCAGTATGCTCT 56
DB 69 ATGGGACCTGGATTGTTGCTGCTGCGAGACGACCTTCAGTATGCTCT 124
XX
RESULT 14
ADQ22977
ID ADQ22977 standard; DNA; 852 BP.
XX
AC ADQ22977;
XX
```

```
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5797.
XX
XX Human soft tissue sarcoma; cytosolic; gene therapy; vaccine; screening; human;
XX de.
XX
OS Homo sapiens.
XX
XX WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnick A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
PS Example 2; SEQ ID NO 5797; 210pp; English.
XX
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 852 BP; 226 A; 267 C; 169 G; 190 T; 0 U; 0 Other;
XX
Query Match 6.4%; Score 43.2; DB 12; Length 852;
Best Local Similarity 85.7%; Pred. No. 0.33;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
QY 1 ATGGGACCTGGATTGTTGCTGCTGCGAGACGACCTTCAGTATGCTCT 56
DB 69 ATGGGACCTGGATTGTTGCTGCTGCGAGACGACCTTCAGTATGCTCT 124
XX
RESULT 15
ABQ88146
ID ABQ88146 standard; cDNA; 17896 BP.
XX
AC ABQ88146;
XX
DT 18-SEP-2002 (first entry)
XX
DE Human osteoblast differentiation related cDNA SEQ ID NO 53.
XX
XX Human; osteoblast; stem cell differentiation; bone tissue deposition;
XX osteoporosis; osteopathic; ss.
XX
OS Homo sapiens.
XX
XX WO200250301-A2.
XX
```

PD 27-JUN-2002.
XX
XX 18-DEC-2001; 2001WO-US048276.
XX
XX 18-DEC-2000; 2000US-0255882P.
XX
XX 24-APR-2001; 2001US-0285691P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PROC) PROCTER & GAMBLE CO.
XX
XX J1 D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
XX Mertz L;
XX MPI; 2002-557663/59.
XX
XX Use of genes and their expression profiles associated with osteoblast
XX differentiation for screening modulators bone formation, for diagnosing
XX or treating e.g. osteoporosis, or as markers for the differentiation
XX process.
XX
XX Claim 1; SEQ ID NO 53; 78bp + Sequence Listing; English.
XX
XX The invention relates to genes and their expression profiles are used
XX for: (a) screening modulators of precursor stem cell differentiation into
XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
XX deposition of bone tissue, abnormal rate of osteoblast formation or
XX osteoporosis; or (c) treating or monitoring treatment of the conditions
XX cited in (b), or monitoring the progression of bone tissue deposition.
XX Specific conditions include postmenopausal osteoporosis, glucocorticoid
XX osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
XX induced abnormalities in bone formation or bone loss, conditions that
XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
XX or fibrous dysplasia. The present sequence is that of an osteoblast
XX differentiation associated CDNA marker of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 178896 BP; 50288 A; 36724 C; 39084 G; 52800 T; 0 U; 0 Other;
XX
XX Query Match 6.3%; Score 43; DB 6; Length 178896;
XX Best Local Similarity 46.5%; Pred. No. 1.8;
XX Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
XX
XX QY 165. CACAAATGTCAGTGTGTTAAGAAACATTTTAAATCTTGATCACAAGATTTTAAACA 224
XX Db 73412 CAAAAATGACAAATGGAATCATCATCAAGTTAAAGCTTCTGCACAGCCAGATACAA 73471
XX
XX QY 225 AAAACATTCACATTCGCTTCAACCATATTCGATTCAGTATAGCCAGAGTGGGGGTG 284
XX Db 73472 CCAACAAAGAGAGCAACCCAGAAATAGAGAAATATCTGTAACTACCATCTG 73531
XX
XX QY 285 GCTGAACACTGTAATTTTAAAGCAAGAACTCAGAAATTTTGAAGCAATTAATT 344
XX Db 73532 ACAAGGATTTAAATACCAAGATATATAGAGCTCAACAACTCTATAGAAAAAATCTA 73591
XX
XX QY 345 ATATATCACTCATCTTAGATGAAATTTTGAAGGACCTTGAAGGCTTCAGAAA 404
XX Db 73592 ATATCTGAATTTAAAAAATGGCAAAAGATTGAATGACATTTCTCAAAAAGAACATA 73651
XX
XX QY 405 AAGTGCTCAATCAGTCTTTAAGTATCAAAAATGCCAGTTTCTTAAACCNAATTT 463
XX Db 73652 CAGAGGCAAAAGAGGATGATATCATGATTAATGAATAATCACAACCTATATATCATAT 73710

Search completed: February 8, 2005, 13:57:24
Job time : 905.049 secs

This Page Blank (uspto)


```
; Sequence 3, Application US/09744128
; Patent No. 6677306
; GENERAL INFORMATION:
; APPLICANT: Veis et al.
; TITLE OF INVENTION: Chondrogenic and Osteogenic Inducing Molecule
; FILE REFERENCE: 27636/36983
; CURRENT APPLICATION NUMBER: US/09/744,128
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: PCT/US99/17342
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: 60/094,489
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn 3.1
; SEQ ID NO 3
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PCR Product
US-09-744-128-3

Query Match
Best Local Similarity 90.6%; Score 45; DB 4; Length 476;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY
1 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 53
48 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 100

RESULT 3
US-09-018-595B-3
; Sequence 3, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Weitz,
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for Windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: PELM-744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-018-595B-1
; Sequence 1, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
```

```
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-018-595B-3

Query Match
Best Local Similarity 87.5%; Score 44.8; DB 2; Length 500;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY
1 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 56
69 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 124

RESULT 4
US-09-324-709A-3
; Sequence 3, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; APPLICANT: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; TITLE OF INVENTION: SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Weitz,
; ADDRESSEE: Wilson Sonsini Goodrich & Rosati
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for Windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weitz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-324-709A-3

Query Match
Best Local Similarity 87.5%; Score 44.8; DB 3; Length 500;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY
1 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 56
69 ATGGGACCTGGATTGTTGCTGCTGCTGCGGAGCAGCCTTCAGTATGCC 124

RESULT 5
US-09-018-595B-1
; Sequence 1, Application US/09018595B
; Patent No. 5962233
; GENERAL INFORMATION:
```



```

; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; TITLE OF INVENTION: AMPLIFICATION PRODUCTS AT MULTIPLE ALLELIC SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: PELM-744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-018-595B-1
;
Query Match 6.6%; Score 44.8; DB 2; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.0052;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTTATTGCTGCTGCTGCGAGCAGCCTTTGCGATGCTCT 124

RESULT 6
US-09-324-709A-1
; Sequence 1, Application US/09324709A
; Patent No. 6154707
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: Applied Biosystems Division
; TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: Wordperfect for windows 6.0,
```

```

; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/324,709A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: 16842-758
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-324-709A-1
;
Query Match 6.6%; Score 44.8; DB 3; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.0052;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTTATTGCTGCTGCTGCGAGCAGCCTTTGCGATGCTCT 124

RESULT 7
US-09-018-595B-2
; Sequence 2, Application US/09018595B
; Patent No. 596223
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: Applied Biosystems Division
; TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weltz,
; STREET: 650 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1050
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Wordperfect for windows 6.0,
; SOFTWARE: ASCII (DOS) TEXT format
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,595B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Weltz
; REGISTRATION NUMBER: 38,362
; REFERENCE/DOCKET NUMBER: PELM-744
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 493-9300
; TELEFAX: (650) 493-6811
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 nucleotides
```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-018-595B-2

Query Match 6.4%; Score 43.2; DB 2; Length 802;
Best Local Similarity 85.7%; Pred. No. 0.015;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTTCCTGCTCGGAGACCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGATTTTGTTCCTGCTCGGAGACCTTCAGTATGCTGT 124

RESULT 8
US-09-324-709A-2
Sequence 2, Application US/09324709A
Patent No. 6154707

GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
APPLICANT: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
TITLE OF INVENTION: SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: David J. Weltz,
ADDRESSEE: Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
SOFTWARE: ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: David J. Weltz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758

TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300

TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

US-09-324-709A-2

Query Match 6.4%; Score 43.2; DB 3; Length 802;
Best Local Similarity 85.7%; Pred. No. 0.015;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGGGACCTGATTTTGTTCCTGCTCGGAGACCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGATTTTGTTCCTGCTCGGAGACCTTCAGTATGCTGT 124

RESULT 9
US-09-806-708B-22/C
Sequence 22, Application US/09806708B
Patent No. 6784342

GENERAL INFORMATION:

APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741

CURRENT APPLICATION NUMBER: US/09/806,708B
CURRENT FILING DATE: 2001-04-03

PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04

NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0

SEQ ID NO 22
LENGTH: 1141

TYPE: DNA
ORGANISM: Artificial sequence

FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1141)

OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEI promoters
US-09-806-708B-22

Query Match 6.0%; Score 40.8; DB 4; Length 1141;
Best Local Similarity 8.0%; Pred. No. 0.079;
Matches 39; Conservative 213; Mismatches 233; Indels 0; Gaps 0;

QY 182 TTAAGAAACATTTTAAATCTTGATCACAAGATTTTAAACAAACATTCAGTGC 241
DB 626 WTAKGDMTVRKVKVKKRDITCTYVDWADSWWVYANMRCDVYTRNNYCKSYAHSY 567

QY 242 CTTCACCATATTCGATTCAGTATAGCCAGATGGGGGTGCTGAGAACCTGGAAT 301
DB 566 WYNSNNAMWYRYSAANWSMARWTRNNMWSGVBWRMAGTMMRNNNNNTDPRYRW 507

QY 302 TTAACAGCAAGAACCTCGAAATTCCTGGAACATTTATATATCATCATCTCTA 361
DB 506 WKRWARBTYYTDDSCNCAKSMWRGNMRAMKMMWANNDAAGMDWTYMGNTYMMER 447

QY 362 GATGAAATTTTGGAGAGGACCTTGAAGGCTCCAGAAAGAGTGTCAATCATGCT 421
DB 446 AMKMMNMACRAAYCCNNNNNPAACVHKKMKRWTKYMKACNNNNBKAMTRVAMM 387

QY 422 TTAAGTACTCAAAAATGACAGTTTCTCTAAACCAATTTCTCAAGATCTAATC 481
DB 386 YSRDTNTDMMWTSDBWHWYTVDYMMRAVNNNNNNNNWBRCKTSMWMMMDHMTCT 327

QY 482 TCTTCTGCTCAAAAACCCCGGCTATCATTTCTGAGATGAAGATTCAGTGTAGA 541
DB 326 YGNTWGSAYBMAAASMBVYVWVCWRTYMGKTMNNNNNNNRAWYRTTYVAM 267

QY 542 ATTTCTCTGCAATTTATTTGTCAGTCACTGCTAAGAGATGATCTCTATTGAT 601
DB 266 CNRRYYDPAVWTRKRNRYCTAYBMYVYMGKHBWRRAHRSMMWVYKCNKYM 207

QY 602 TGTCCAGATGATGATGATGATGATTTACATTCAAATGATCAAGGATGAGGAATGTG 661
DB 206 VSMHYAMRYBKWABAVGCNNNMKDBAHHCATNNNNMMWYAVHHMKKGAAMTN 147

QY 662 GTTAC 666
DB 146 NKTAB 142

RESULT 10
US-09-248-796A-11112
Sequence 11112, Application US/09248796A

Patent No. 6747137
GENERAL INFORMATION:

APPLICANT: Keith Weltsch et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAT

FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A

CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725

Query Match	5.7%	Score 38.6;	DB 4;	Length 2478;
Best Local Similarity	46.5%	Pred. No. 0.45;		
Matches 125, Conservative	0;	Mismatches 144;	Indels 0;	Gaps

```

RESULT 13
US-09-328-352-1803
; Sequence 1803, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1803
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-1803

```

Query Match	5.5%	Score 37.4	DB 4	Length 270
Best Local Similarity	55.9%	Pred. No. 0.39		
Matches 71	Conservative 0	Mismatches 56	Indels 0	Gaps 0

QY	109	CAGATTTTATTTAAGGTTAAAATTTAAGGTTTAAAGACAGTACAAGATCTGATCTTACA	168
Db	13	CAAAATTTTATTTTAAATTCATTTAATTTTTTATTTTAAAGATGAGTTAATTATATATA	72
QY	169	AATGTGACTGTGTTTAAGAAACATTTTAAAACTTGATCAACAGATTTTAAAACAAAA	228
Db	73	TTTGAAATTTGTGATTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	132
QY	229	CATTCTC	235
Db	133	AATTATC	139

RESULT 14
US-09-248-796A-4289/C
; Sequence 4289, Application US/09248796A

Query Match	5.5%	Score 37.2;	DB 4;	Length 927;
Best Local Similarity	50.6%;	Pred. No. 0.74;		
Matches 90;	Conservative 0;	Mismatches 88;	Indels 0;	Gaps 0;

QY	89	TTGGAATATAAAATCTGCCTCAGATTTTATTATAGGTTAAATTGAAGGTTTAAAGCAG	148
Db	372	TTGGCTTTTATGATCTTCCCAAGATTTTTTTCTTACCAGATGTTTGAAGATATGATTTAA	313
QY	149	TACAAGATCTGATGTTTACAAAATGAGTCTGTTTAAAGAAACATTTTAAATCTTGATC	208
Db	312	CCCAAAATGCTCTGTGACATATGCGCTCTGAGAGAAAACATGAGCAACTCAAAATC	253
QY	209	ACAAAGTTTTAAACAAAACATTTCTCAGTGCCTTACCCCATATTTCTGATTTAGAT	266
Db	252	CCAAGGATACTTTAAATTTCCATTCATCTCTTGATGTTGATATTTTATCCACTTT	195

```

RESULT 15
US-09-134-001C-2435
; Sequence 2435, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

```

```

; SEQ ID NO 2435
;
; LENGTH: 357
;
; TYPE: DNA
;
; ORGANISM: Staphylococcus epidermidis
;
US-09-134-001C-2435

```

Query Match	5.4%	Score 36.8;	DB 3;	Length 357;
Best Local Similarity	53.5%;	Pred. No. 0.64;		
Matches 77;	Conservative 0;	Mismatches 67;	Indels 0;	Gaps 0

QY	87	GCTTGGAAATTAATAAACTGCTCCACATTTTATTAATGGGTTAAATTTAAAGGGTTTAAAGAC	146
QY			
Db	118	GGTTGGAAGTTTGGAAGATGCAATTAAGACCTCTTTGGAGATGAACGTAAGTAATGGAT	177
QY	147	AGTACAGAATTCGATGTTTACAAATAGTAGCTGTGTTTAAAGAAATTTTAAATCTTGA	206
QY			
Db	178	AGTTTAAACCTTAAGAGAGCGGAATTCGTATTAAAGAGAAATGGCACAACGTAGA	237
QY	207	TCCAAGATTTTAAACCAAAACA	230
QY			
Db	238	TTCAAGAAGTTGAACGTAAAGAAA	261

Search completed: February 8, 2005, 22:26:23
Job time : 185.767 secs

QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACGATCAAGATCGATGTTCA 168
DB 69991 -----TCATAGTTGGTGAATATGAGTTTAAACAGTATGAGATTCATTTCA 70041
QY 169 AATGTACTGTGTGTTAAGAAACATTTTAAATCTTGATCAAGATTTTAAAAA 228
DB 70042 TAGTCTCTGGGTGAAGAAACCTTCAGAGCTT-----GTTTAAAAAGTA 70090
QY 229 CATTCCTAGTGTGCTTACCCATATCTGATTCAGATAGCCAGATGGGGGTGCTG 288
DB 70091 TATTCCTCAATGCGCTTACCAAAA-----ATTCTGATTTGGTACAGCTGGGGGGGGCC 70145
QY 289 AGAAGCTGAAATTTTAAACAACA-----AGAAGCTGAAATTTCTGAAACAATTA 344
DB 70146 AGGACTCTGATTTTAAAGCAACCCAGAGATTTCTTTGGAAGCTTATGCTTGAAT 70205
QY 345 ATAACTACTCATCTCTAGATGA--AAATTTTGAAGGAGCCTTTGAAAGCCTTCAGA 402
DB 70206 ATCAACACCCATCTCTAGATGAGAAAGCTTTTGAAGGAGCCTTTGAAAGCTTCAGA 70265
QY 403 AAAAGCTCATCATCTCTT---AGTACTAAGAAATGCGAGTTTCTTAAACCA 459
DB 70266 GAAAGGCTTAACACCTTGTGACCAATATTAAGAGATGCGAGTTTGTCTAAACCA 70325
QY 460 ATTCTCACAAGATTAATCTCTCTTCTGCTCAGAAACCCCGGCTATCATCTTGA 519
DB 70326 ATTCCTCTCAAGATTTCAATCTCTCTGCTTCCATATGCTGCTTATCCCTCA 70385
QY 520 --GAGTGAATTTCAAGTGTGAGAAATTTCTTCTGAACTATTATTGT 567
DB 70386 GGGGGTGAATTTTGTGTAGGAATTCACCTTTTGAAGCACTTCTCTGT 70435

RESULT 2
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050

Query Match 11.8%; Score 80.4; DB 13; Length 600;
Best Local Similarity 70.4%; Pred. No. 1.3e-10;
Matches 145; Conservative 1; Mismatches 37; Indels 23; Gaps 2;
QY 1 ATGGGAGCTGTGATTTTGTGCTGCTCTGCGAGAGAGCCTTCAGATGCTGTGAGT 60
DB 256 ATGGGAGCTGTGATTTTGTGCTGCTCTGCGAGAGAGCCTTCAGATGCTGTGAGT 197

QY 61 AAAATTTT-----CAATTTCAATTTCAAGCTTGAATTAATCTGCT 108
DB 196 AAAACACCCCTTGCAATAGCTGATGTCATTTCAAACTTGACATTAATCTGCT- 138
QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACGATCAAGATCTGATGTTCA 168
DB 137 -----CATAGTTGGTGAATTTAAGGTTTAAACAGTATGAGATCAGATGCTTCA 88
QY 169 AATGTACTGTGTGTTAAGAAACATTT 194
DB 87 TAGTCTCTGGGTGAAGAAACACTT 62

RESULT 3
US-10-027-632-287050/c
; Sequence 287050, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287050
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-287050

Query Match 11.8%; Score 80.4; DB 15; Length 600;
Best Local Similarity 70.4%; Pred. No. 1.3e-10;
Matches 145; Conservative 1; Mismatches 37; Indels 23; Gaps 2;
QY 1 ATGGGAGCTGTGATTTTGTGCTGCTCTGCGAGAGAGCCTTCAGATGCTGTGAGT 60
DB 256 ATGGGAGCTGTGATTTTGTGCTGCTCTGCGAGAGAGCCTTCAGATGCTGTGAGT 197
QY 61 AAAATTTT-----CAATTTCAATTTCAAGCTTGAATTAATTAATCTGCT 108
DB 196 AAAACACCCCTTGCAATAGCTGATGTCATTTCAAACTTGACATTAATCTGCT- 138
QY 109 CAGATTTTATTAGGTTAAATTTAAGGTTTAAAGACGATCAAGATCTGATGTTCA 168
DB 137 -----CATAGTTGGTGAATTTAAGGTTTAAACAGTATGAGATCAGATGCTTCA 88
QY 169 AATGTACTGTGTGTTAAGAAACATTT 194
DB 87 TAGTCTCTGGGTGAAGAAACACTT 62

RESULT 4
US-10-027-632-287049/c
; Sequence 287049, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:

```

/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027.632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 287049
/ LENGTH: 600
/ TYPE: DNA
/ ORGANISM: Human
/ US-10-027-632-287049

```

Query Match	11.6%	Score 79.2;	DB 13;	Length 600;
Best Local Similarity	66.2%	Pred. No. 2.6e-10;		
Matches 129; Conservative	0;	Mismatches 65;	Indels 1;	Gaps 1;

Qy 1 ATGGGGACCTGGAGATTTTGGTTTGGCTGCTCCCTCGGGAGAGAGCTTCAGATATCCCTGAGT 60

Db 256 ATGGGGACCTGGAGATTTTATTTGGCTGCTCCCTCGGGAGAGAGCTTTTGGCATATCCTGTAGT 197

Qy 61 AAAATTTCCAAATTTCCAAATTTCAAGAGCTTGGAAAAATAAATATGCTCCAGATTTTAT- 119

Db 196 AAAACACCCCTTCATATAGTCAGTGTCCAAATTTCAAAATCTGGACATATAAAATCTGTCTC 137

Qy 120 TTTAGGGTTAAATTTAAGGGTTTAAAGACAGTACAGATCTGATGTTCAAAATGTAAGTCT 179

Db 136 ATAGTTGTGTAATTTAGGGTTTAAAAAGATATGAGATCATATGTCCTCTGG 77

Qy 180 GTTAAAGAAACATT 194

Db 76 GTTGAAGAAACACTT 62

```

RESULT 5 -
US-10-027-613-287049/c
/ Sequence 287049, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108927.129
/ CURRENT APPLICATION NUMBER: US/10/027.632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002

```

```

: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 325720
: SOFTWARE: FASTSEQ For Windows Version 4.0.
: SEQ ID NO 287049
: LENGTH: 600
: TYPE: DNA
: ORGANISM: Human
:
US-10-027-632-287049

```

Query Match	11.6%;	Score 79.2;	DB 15;	Length 600;
Best Local Similarity	66.2%;	Pred. No. 2.6e-10;		
Matches 129;	Conservative 0;	Mismatches 65;	Indels 1;	Gaps 1;

QY	1	ATGGGGACCTGGAGTTTGGTTTGGTCCGACCTCCGGAGAGAGCTTCAGATATCCCTGTAGT	60
Db	256	ATGGGAGCTGGATTTTATTTTGGTCCGTCCCTCTGGAGAGAGCTTTTGGCATGCTGTGAGT	19
QY	61	AAATTTCCAAATTTCCAAATTTCCACAAGCTGGAAATAAAAATCTGCTCAGATTTTAT-	115
Db	196	AAAACACCCCTTGCATTAAGTCAGTGTCCAAATTTGCAAACTTGGACATTAATAATCTGCTC	13
QY	120	TTAGGCTTAAATTTAAGGCTTTTAAGACAGTACAGATCTGATGTTCCAAATGTGACTGT	175
Db	136	ATAGTTGTGTAATTTAGGCTTTTAAAAACAGTATGAGATCAGATGTCCTTCATATGCTCTGG	77
QY	180	GTTTAAGAACATTT	194
Db	76	GTTGAAGAAACCTT	62

RESULT 6
US-10-104-774-3

```

: APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
:
: TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE
:
: TITLE OF INVENTION: SITES
:
: FILE REFERENCE: 16842-782
:
: CURRENT APPLICATION NUMBER: US/10/104,774
:
: CURRENT FILING DATE: 2002-03-21
:
: PRIOR APPLICATION NUMBER: US 09/018,595
:
: PRIOR FILING DATE: 1998-02-04
:
: NUMBER OF SEQ ID NOS: 11
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 3
:
: LENGTH: 500
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: OS-10-104-774-3

```

Query Match	6.6%	Score 44.8;	DB 13;	Length 500;
Best Local Similarity	87.5%;	Pred. No. 0.25;		
Matches 49;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;

Dy 1 ATGGGACCTGGATTTTGTTCCTGCCCTCGGAGCAGCTTCAGTATGCTGT 56

Db 69 ATGGGACCTGGATTTTGTTCCTGCCCTCGGAGCAGCTTTTGCCATGCTCT 124

RESULT 7
US-10-455-150-3
; Sequence 3, Application US/10455150
; Publication No. US20040053302A1
; GENERAL INFORMATION:
; APPLICANT: Perkin-Elmer Corporation,
; TITLE OF INVENTION: Applied Biosystems Division
; FLOURESCENCE GENOTYPING AT MULTIPLE ALLELIC
; SITES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Weitz,
Wilson Sonsini Goodrich & Rosati

STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-10-455-150-3
Query Match 6.6%; Score 44.8; DB 16; Length 500;
Best Local Similarity 87.5%; Pred. No. 0.25;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 124

RESULT 8
US-10-104-774-1
Sequence 1, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
TITLE OF INVENTION: SITES
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 09/018,595
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-1

Query Match 6.6%; Score 44.8; DB 13; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.31;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 124

RESULT 9
US-10-455-150-1
Sequence 1, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:
APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
TITLE OF INVENTION: FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Weitz,
Wilson Sonsini Goodrich & Rosati
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
ASCII (DOS) TEXT format
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-455-150-1
Query Match 6.6%; Score 44.8; DB 16; Length 793;
Best Local Similarity 87.5%; Pred. No. 0.31;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 56
DB 69 ATGGGACCTGGATTGTTGCTGCTGCTGCGAGCAGCCTTCAGTATGCTGT 124

RESULT 10
US-10-755-889-605
Sequence 605, Application US/10755889
Publication No. US20040171823A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
PATHWAY
FILE REFERENCE: D0284 NP
CURRENT APPLICATION NUMBER: US/10/755,889
CURRENT FILING DATE: 2004-01-13
PRIOR APPLICATION NUMBER: U.S. 60/440,068
PRIOR FILING DATE: 2003-01-14
PRIOR APPLICATION NUMBER: U.S. 60/469,757
PRIOR FILING DATE: 2003-05-12
NUMBER OF SEQ ID NOS: 823
SOFTWARE: PatentIn version 3.2

SEQ ID NO 605
LENGTH: 793
TYPE: DNA
ORGANISM: Homo sapiens
US-10-755-889-605

Query Match
Best Local Similarity 6.4%; Score 44.8; DB 17; Length 793;
Matches 49; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 124

RESULT 11
US-10-104-774-2
Sequence 2, Application US/10104774
Publication No. US20020164630A1
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation, Applied Biosystems Division
TITLE OF INVENTION: DETERMINATION OF GENOTYPE OF AMPLIFICATION PRODUCTS AT MULTIPLE A
FILE REFERENCE: 16842-782
CURRENT APPLICATION NUMBER: US/10/104,774
CURRENT FILING DATE: 2002-03-21
PRIOR FILING DATE: 1998-02-04
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 802
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-774-2

Query Match
Best Local Similarity 6.4%; Score 43.2; DB 13; Length 802;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 124

RESULT 12
US-10-455-150-2

Sequence 2, Application US/10455150
Publication No. US20040053302A1
GENERAL INFORMATION:

APPLICANT: Perkin-Elmer Corporation,
TITLE OF INVENTION: Applied Biosystems Division
FLUORESCENCE GENOTYPING AT MULTIPLE ALLELIC
SITES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSES: David J. Weitz,
STREET: 650 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1050
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Wordperfect for windows 6.0,
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/455,150
FILING DATE: 04-Jun-2003
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/324,709A
FILING DATE: 03-June-1999
ATTORNEY/AGENT INFORMATION:
NAME: David J. Weitz
REGISTRATION NUMBER: 38,362
REFERENCE/DOCKET NUMBER: 16842-758
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 493-9300
TELEFAX: (650) 493-6811
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-455-150-2

Query Match
Best Local Similarity 6.4%; Score 43.2; DB 16; Length 802;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 124

RESULT 13
US-10-723-860-5797

Sequence 5797, Application US/10723860
Publication No. US20040253606A1
GENERAL INFORMATION:

APPLICANT: Aziz, Natasha
APPLICANT: Ginsburg, Wendy M.
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
FILE REFERENCE: 05882.0193.NPUS01
CURRENT FILING DATE: 2003-11-26
PRIOR APPLICATION NUMBER: 60/429,739
PRIOR FILING DATE: 2002-11-26
NUMBER OF SEQ ID NOS: 8393
SOFTWARE: PatentIn version 3.2
SEQ ID NO 5797
LENGTH: 852
TYPE: DNA
ORGANISM: Homo sapiens
US-10-723-860-5797

Query Match
Best Local Similarity 6.4%; Score 43.2; DB 18; Length 852;
Matches 48; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 56
Db 69 ATGGGACCTGATTTGTTGCTGCTGCTGCGAGACAGCTTCAGTATGCTGT 124

RESULT 14
US-10-450-826-53

Sequence 53, Application US/10450826
Publication No. US20040101818A1
GENERAL INFORMATION:

APPLICANT: JI, Darren
APPLICANT: Cook, Jonathan S.
APPLICANT: Uatawal, Neelam
APPLICANT: Bistein, Richard
APPLICANT: Houghton, Adam
APPLICANT: Weitz, Lawrence
TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiat

FILE REFERENCE: 044921-5039-WO
CURRENT APPLICATION NUMBER: US/10/450,826
CURRENT FILING DATE: 2003-06-18
PRIOR APPLICATION NUMBER: US 60/255,882
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: US 60/285,691
PRIOR FILING DATE: 2001-04-24
NUMBER OF SEQ ID NOS: 149
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 53
LENGTH: 178896
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. AC008041
US-10-450-826-53

Query Match 6.3%; Score 43; DB 17; Length 178896;
Best Local Similarity 46.5%; Pred. No. 14;
Matches 139; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 165 CAGAAATGACGCTGTTTAAAGAAACATTTTAAATCTTGATCAGAGATTTTAAACA 224
DB 73412 CAAAAATGACAAATGAAATCAATCAAGTTAAAGCTTCTGCACAGCCAGATACAA 73471
QY 225 AAAACATTCCTAGTTCCTTACCCATATTCGATTCAGTATAGCCAGAGTGGGGGTGT 284
DB 73472 CCAACAAAGAGAGAGCAACCCACAGATAGAGAAATATCTGTAATCTACCATCTG 73531
QY 285 GCTGAGAACTGTAATTTTAAAGCAAGAACTCAGAAATCTTGAAACATTAATT 344
DB 73532 ACAGAGATTATATACCAATATATAGAGCTCAACAACTCTATAGAAAAATCTA 73591
QY 345 ATATCACTATCTCTAGATGAAATTTTGAAGGACCTTGAAGGCTCCAGAAA 404
DB 73592 ATATCTGATTTTAAAAATGGGCAAAAGATTGATAGACATTTCTCAAAAGAAACATA 73651
QY 405 AAGTCTCAATCACTCTTTTAAAGTAACTAACAAAATGCCAGTTTCTTAAACCAATTT 463
DB 73652 CAGAGGGCAAAACAGCATTTGATATCAATTAATGAATCAACAACTATATCATATAT 73710

RESULT 15
US-10-417-375-66
Sequence 66, Application US/10417375
Publication No. US20040219528A1
GENERAL INFORMATION:
APPLICANT: David W. Morris
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
FILE REFERENCE: 529452001600
CURRENT APPLICATION NUMBER: US/10/417,375
CURRENT FILING DATE: 2003-04-15
NUMBER OF SEQ ID NOS: 176
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 252907
TYPE: DNA
ORGANISM: Homo sapiens
US-10-417-375-66

Query Match 6.2%; Score 42.4; DB 18; Length 252907;
Best Local Similarity 52.2%; Pred. No. 24;
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 76 CAATTCACAGCTTGAAATTAATAATCTGCTCAGATTTTATTTAGGGTTAAATTTAA 135
DB 177651 CACATGAAATCTTTTAAACAAATGGAATCACTTTTATTAATTTGATTTT 177710
QY 136 GGGTTTAAAGACGTAAGATGATGTTTCAAAATGAGCTGTTTAAAGAAATTT 195
DB 177711 AAATTTAAAGATATATGAAATATTTTCCATATTTGATGATATATATATATAT 177770

QY 196 TAAATCTGATCAGAGATTTTAAACAAACATTTCTGAGTGCCTGCACCATATTC 255
DB 177771 TATAATATGATGTTATGATATATATATATATTTTGAAGATGAGCTCACTTGT 177830
Search completed: February 9, 2005, 06:51:29
Job time : 925.298 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 8, 2005, 12:58:25 ; Search time 8099.44 Seconds
(without alignments)
3059.349 Million cell updates/sec

Title: US-10-754-437-23

Perfect score: 680
Sequence: 1 atggggagactgagatttgtt.....ggtactgagagacagacac 680

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	90.6	13.3	516	8	AQ003910 CIT-HSP-2
2	49.6	7.3	429	6	CB473334 sm68_E03.
3	46.6	6.9	303	7	W33932 mb54d02.r1
4	46.6	6.9	430	7	W40649 mb43h12.r1
5	46.6	6.9	536	7	W36345 mb72c12.r1
6	46.6	6.9	585	9	AY419441 Mus muscu
7	46.6	6.9	645	6	CD773419 AGENCOURT
8	46.6	6.9	707	6	CB056709 NISC_j119
9	46.6	6.9	812	3	AK029358 Mus muscu
10	46.6	6.9	843	6	CB588212 AGENCOURT
11	46.6	6.9	859	6	CB588525 AGENCOURT
12	46.6	6.9	865	6	CB589177 AGENCOURT
13	46.6	6.9	886	6	CB587051 AGENCOURT
14	46.6	6.9	889	6	CB589251 AGENCOURT
15	46.6	6.9	891	6	CB574837 AGENCOURT
16	46.6	6.9	913	6	CB590451 AGENCOURT
17	46.6	6.9	919	6	CB587332 AGENCOURT
18	46.6	6.9	928	6	CB590111 AGENCOURT
19	46.6	6.9	945	2	BB614068
20	46.6	6.9	990	7	W12906
21	46.6	6.9	1006	7	W29475
22	46.6	6.9	1020	7	W08102
23	46.6	6.9	1101	9	CNS0039G
24	45.6	6.7	293	7	R47143

25	45.6	6.7	556	7	R46934	R46934 Y169 Rat in
26	45	6.6	217	7	R46913	R46913 Y140 Rat in
27	45	6.6	223	2	BB571643	BB571643 BB571643
28	45	6.6	395	7	R46955	R46955 Y195 Rat in
29	45	6.6	449	7	R47024	R47024 Y350 Rat in
30	45	6.6	1031	6	CB575508	CB575508 AGENCOURT
31	44.8	6.6	529	9	CR342574	CR342574 Medicago
32	44.8	6.6	570	9	AY419439	AY419439 Homo sapi
33	44.8	6.6	570	9	AY419440	AY419440 Pan trogl
34	44.6	6.6	303	7	R46903	R46903 Y124 Rat in
35	44.6	6.6	318	7	R47078	R47078 Y534 Rat in
36	44.6	6.6	467	7	R46947	R46947 Y186 Rat in
37	44.4	6.5	987	9	CNS014PQ	AL104456 Drosophila
38	44.4	6.5	373	7	R47135	R47135 Y722 Rat in
39	44	6.5	395	7	R47030	R47030 Y359 Rat in
40	43.8	6.4	1101	9	CNS00386	AL064084 Drosophila
41	43.8	6.4	1313	7	CK997149	CK997149 ip15C02.b
42	43.4	6.4	257	7	R47100	R47100 Y601 Rat in
43	43.2	6.4	727	8	BH705230	BH705230 BOHVM06TF
44	43.2	6.4	783	4	BG188114	BG188114 RST17499
45	43.2	6.4	1101	9	CNS0039G	AL063921 Drosophila

ALIGNMENTS

RESULT 1
LOCUS AQ003910/c 516 bp DNA linear GSS 26-JUN-1998
DEFINITION CIT-HSP-2289L4.TF CIT-HSP Homo sapiens genomic clone 2289L4,
genomic survey sequence.
ACCESSION AQ003910
VERSION AQ003910.1 GI:3081561
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 516)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.B., Basse,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2289L4.TR
Contact: Mark Adams
Department of Bukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.
FEATURES
source Location/Qualifiers
1..516
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7150375"
/db_xref="taxon:9606"
/clone="2289L4"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelBAC11, Site_1: HindIII, Site_2:
HindIII"
ORIGIN
Query Match 13.3%; Score 90.6; DB 8; Length 516;

Best Local Similarity 69.0%; Pred. No. 7,4e-12;
Matches 169; Conservative 0; Mismatches 69; Indels 7; Gaps 3;

QY 330 TGAACAATTACTTATTAATCACTCATCTCTAGATG--GAAATTTTGGAGGACCTT 387
DB 493 TGTAGCTTGTAAATATCACACCCATCTATAGATGAAGAAGGCTTTTGGAGGACCTT 434
QY 388 TGAAGGCTCCGAGAAAGTGTCAATCACTCTCT--TAAGTACTACAAAATGTCAGT 444
DB 433 TGAAGGCTCCGAGAAAGTGTCAATCACTCTCTCTTGGACAAATATGCAAGATGCCAGT 374
QY 445 TTTCTCTAAAACCAATTTCTCAAGATATCACTCTCTTCTGCTCAGAAAACCCCG 504
DB 373 TTTGTCTAAAACCAATTTCTCAAGATATCACTCTCTCTGCTCAGAAATGTC 314
QY 505 GCGTATCATCTTTCA--GAGTAAAGATTTCAGTGTAGAAATTTCTCTGAACTATTA 562
DB 313 TGTCTTACCCCTCAGGGGCTAAGATTTTGTGTAGAAATCACTTTTGGACCCACAT 254
QY 563 TTTGT 567
DB 253 CCTGT 249

RESULT 2
CB473334 429 bp mRNA linear EST 26-MAR-2003
LOCUS CB473334
DEFINITION B68_E03.f sn Sus scrofa CDNA 5', mRNA sequence.
ACCESSION CB473334
VERSION CB473334.1 GI:29279720
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE
AUTHORS Bakaryota; Metzosa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 429)
Neilan,D.G., Kutish,G.F., Lu,Z., Zeak,A. and Rock,D.L.
Sequence analysis of African swine fever virus infected and
non-infected porcine macrophage CDNA libraries
Unpublished (2003)

JOURNAL
COMMENT Contact: Neilan JG
Plum Island Animal Disease Center
US Department of Agriculture, Agricultural Research Service
PO Box 848, Greenport, NY 11944-848, USA
Tel: 631 323 3133
Fax: 631 323 3044
Email: jneilan@piadc.ars.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim.alg option. Vector identified by
cross match v0.990329 and Lucy v1.17p.
Seq primer: M13 Forward.

FEATURES
Source Location/Qualifiers

1..429
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="lymphoid"
/cell_type="macrophage"
/lab_host="DH10B"
/clone_lib="sn"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;
Library made from pools of polyA selected RNA, isolated at
different times post-infection (0 to 16 hrs) from African
swine fever virus (isolate Pretoriuskop/96/4) infected
swine macrophages. Macrophages were derived from
peripheral blood mononuclear cells cultured for 48 hrs on
plastic in the presence of 30% F129 supernatant."

ORIGIN

Query Match 7.3%; Score 49.6; DB 6; Length 429;
Best Local Similarity 92.9%; Pred. No. 0.13;
Matches 52; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGCAGCCTTCAGTATGCCGT 56
DB 368 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGCAGCCTTCAGTATGCCCT 313

RESULT 3
W33932 303 bp mRNA linear EST 11-SEP-1996
LOCUS mb54d02.r1 Soares mouse p3NMFI9.5 Mus musculus CDNA clone
DEFINITION IMAGE:333219.5' similar to gb:U10095 Mouse amelogenin (MOUSE);
mRNA sequence.

ACCESSION W33932
VERSION W33932.1 GI:1315837
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Bakaryota; Metzosa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303)
Marra,M., Hillier,L., Allen,M., Bowler,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,P., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
WGI:214619
Seq primer: mob.REGA+ET
High quality sequence stop: 185.

FEATURES
Source Location/Qualifiers

1..303
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:333219"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMFI9.5"
/note="Vector: p773D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand CDNA
was primed with a Not I - oligo(dT) primer (5'
TGTATCCATCTGAGTGGAGCGGCGCATTTTGTGCTGCTCTGGAGCAGCCTTCAGTATGCC 3'),
double-stranded CDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 6.9%; Score 46.6; DB 7; Length 303;
Best Local Similarity 92.5%; Pred. No. 0.71;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGCAGCCTTCAGTATGCC 53
DB 70 ATGGGACCTGGATTTTGTGCTGCTCTCTGGAGCAGCCTTCAGTATGCC 122

RESULT 4
W40649 430 bp mRNA linear EST 11-SEP-1996
LOCUS mc43h12.r1 Soares mouse p3NMFI9.5 Mus musculus CDNA clone
DEFINITION

ACCESSION	IMAGE:351335_5' similar to gb:M10095 Mouse amelogenin (MOUSE); , mus sequence.
VERSION	W40649
KEYWORDS	WT0649.1 GI:1324982
SOURCE	Mus musculus
ORGANISM	Mus musculus (house mouse)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus. 1 (bases 1 to 430)
AUTHORS	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouq, T., Geisler, S., Kucaba, T., Lacy, M., Martin, J., Morris, M., Schellenberg, R., Stepien, M., Tan, F., Underwood, K., Moore, B., Thieling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LINDA; contact the IMAGE Consortium (info@image.lind.gov) for further information. MG1:223135
FEATURES	Seq primer: ESTprimer High quality sequence stop: 323. Location/Qualifiers 1..430
SOURCE	/organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE:351335" /dev_stage="19.5 dpc total fetus" /lab_host="DH10B (ampicillin resistant)" /clone_lib="Soares mouse p3MMF19.5" /note="Vector: p7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5]. TGTTACCAATCGAGGAGGGAGCGGCCCATTTTTTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Patima Bonaldi. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."
ORIGIN	
Query Match	6.9%; Score 46.6; DB 7; Length 430;
Best Local Similarity	92.5%; Pred. No. 0.75;
Matches	49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Dy	1 ATGGGACCTGGATTGTTCCTGCCTCCTGGAGAGACCCTTGATATGCC 53 58 ATGGGACCTGGATTGTTCCTGCCTCCTGGAGAGACCCTTGATATGCC 110
RESULT 5	
LOCUS	W36345 536 bp mRNA linear EST 11-SEP-1996
DEFINITION	mbJ2c12.r1 Soares mouse p3MMF19.5 Mus musculus cDNA clone IMAGE:34966_5' similar to gb:M10095 Mouse amelogenin (MOUSE); , mRNA sequence.
ACCESSION	W36345
VERSION	W36345.1 GI:1318120
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 536)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE
JOURNAL
COMMENT

The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:216366

FEATURES
SOURCE

Seq primer: ETPRimer
High quality sequence stop: 359.
Location/Qualifiers
..536
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:334966"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse pJNMF19.5"
/note="Vector: pT73D (Pharmacia), digested with Not I and cloned into
polylinker; site_1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - Oligo(dt) primer [5'
TGTTACCAATCTGAGAGGGAGCGCCGACTTTTGTATTGTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 6.9%; Score 46.6; DB 7; Length 536;
Best Local Similarity 92.5%; Pred. No. 0.78;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY
DB

1 ATGGGAGACCTGGATTGTTGGCTGCCTCGGAGAGCCTTCAGTATGCC 53
|||||
53 ATGGGAGACCTGGATTGTTGGCTGCCTCGGAGAGCCTTTGCTATGCC 105
|||||

RESULT 6
AY19441

LOCUS
DEFINITION

AY19441 585 bp DNA linear GSS 17-DEC-2003
Mus musculus AMELX gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
AY19441
AY19441.1 GI:39775398

ACCESSION
VERSION
KEYWORDS
SOURCE

GSS.
Mus musculus (house mouse)
Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.

ORGANISM

REFERENCE
AUTHORS

Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Janenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smitsky,D.J.,
Adams,M.D. and Cargill,M.
Interfering nonneutral evolution from human-chimp-mouse orthologous
gene trios

TITLE

JOURNAL
PUBMED
REFERENCE

Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 585)

AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejaritwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriere, S., Wang, G., Zheng, X.H., White, T.J., Shtinsky, O.J., Adams, M.D., and Cargill, M.

TITLE Direct SubMISSION

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA

COMMENT These sequences were made by sequencing genomic exons and ordering them based on alignment.

FEATURES Location/Qualifiers

source 1..585

gene /organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..585
/gene="AMELX"
/locus_tag="HOM6891"

ORIGIN

Query Match 6.9%; Score 46.6; DB 9; Length 585;
Best Local Similarity 92.5%; Pred. No. 0.79; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53
1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53

Db 1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53

RESULT 7 CD773419 645 bp mRNA linear EST 02-JUL-2003

LOCUS AGENCOURT 14713373 NIH MGC 190 Mus musculus cDNA clone

DEFINITION IMAGE:30501267 5', mRNA sequence.

ACCESSION CD773419 GI:32431921

VERSION EST.

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 645)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-r@mail.nih.gov

TITLE Tissue Procurement: Yoshitoko Yamada, Takashi Nakamura, NIDCR

JOURNAL CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

COMMENT DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDCM23 row: k column: 04
High quality sequence stop: 611.
Location/Qualifiers

FEATURES 1..645

source /organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30501267"
/lab_host="DH10B (phage-resistant)"
/lab_host="NIH MGC 190"
/note="Organ: Pooled - Molar; Vector: pDNR-LIB, Site 1:
SfiI (ggccatcagcc); Site 2: SfiI (ggccgctcgcc);
Non-normalized full-length enriched library 5' and 3',
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTAAGCC-3', and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGCGAGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.71

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 645;
Best Local Similarity 92.5%; Pred. No. 0.81; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53
1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53

Db 73 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 125

RESULT 8 CB056709 707 bp mRNA linear EST 17-JAN-2003

LOCUS NISC J19905.w1 Soares NMBP13-15 Mus musculus cDNA clone

DEFINITION IMAGE:4848584 5', mRNA sequence.

ACCESSION CB056709

VERSION CB056709.1 GI:27794996

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 707)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Email: cgabs-r@mail.nih.gov

JOURNAL CDNA Library Preparation:
DNA Sequencing by: The I.M.A.G.E. Consortium/LNL
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
info@image.lnl.gov
Plate: LLM10779 row: N column: 9
Seq primer: T7 primer.
Location/Qualifiers

FEATURES 1..707

source /organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4848584"
/issue_type="pituitary gland"
/dev stage="juvenile, 13-15 days"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares NMBP13-15"
/note="Organ: Brain; Vector: pT73D-Paci, Site 1: NotI,
Site 2: BclRI; 1st strand cDNA was primed with a NotI-
oligo(dT) primer
5'-AACTGGAGAAATTCGGCCGCGGTACGATGTTTTTTTTTTTTTT-3';
double-stranded cDNA was ligated to BclRI adaptors
5'-AATTCGCGCAGAG-3' and 5'-CTGTGCGC-3' (Pharmacia),
digested with NotI and cloned into the NotI and BclRI
sites of the pT73D-Paci vector. Library went through one
round of normalization, and was constructed in the
laboratory of M. Bento Soares (University of Iowa)."

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 707;
Best Local Similarity 92.5%; Pred. No. 0.82; Indels 0; Gaps 0;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53
1 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 53

Db 47 ATGGGACCTGATTTGTTGCTGCTCTGCGAGACCTTCAGTATGCC 99

kb (range 0.5-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA) Corp."

[illegible]

Enyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:<http://genome.gsc.riken.jp/>
URL:<http://fantom.gsc.riken.jp/>

```

FEATURES
    source
        location/Qualifiers
            1..812
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM DB:4833404E21"
                /db_xref="taxon:10090"
                /clone="4833404E21"
                /tissue_type="head"
                /clone_1lb="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="0 day neonate"
            73..663
                /note="unnamed protein product; amelogenin (MGD|MG1:88005, GB|D131768, evidence: BLASTN, 99%, match=799) putative"
                /codon_start=1
                /protein_id="BAC26415.1"
                /db_xref="GI:26325322"
                /translation="MTGWTILFACLLGAAPFLPPHDSFGYINLSYEVLPLKMYOS MIRPDPYSYGVEPMWGQNIIPHSOHPSTLQPRHHLLPVPAQOPVAPAPOOPMM PVPEHSMPTLOTHQNPNIIPSAOPPOPOPOLIPROSHQPMQPGSLPHMQLAQC PLRLPFPSMQPLSLPLELPLEAMPATDKTKREVD"
                793..798
                    /note="putative"
                    812
                        /note="putative"
                                polyA_signal
                                    polyA_site
                                        ORIGIN
                                            Query Match          5.9%; Score 46.6; DB 3; Length 812;
                                            Best Local Similarity 92.5%; Pred. No.0.84;
                                            Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
                                Oy
                                    1 ATGGGAGCACTGATTGTGTTCCTGCCTCCCTGGAGACAGCCTTCAAGTATGCC 53
                                      |||||
                                Db
                                    73 ATGGGAGCACTGATTGTGTTCCTGCCTCCCTGGAGACAGCTTTTGCTATGCC 125
                                      |||||

RESULT 10
LOCUS      CB588212               843 bp     mRNA       linear   EST 03-APR-2003
DEFINITION AGENCOURT 12771359 NIH_MGC_116 Mus musculus cDNA clone
IMAGE      30253239 5', mRNA sequence.
ACCESSION  CB588212
VERSION    CB588212.1  GI:29506068
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 843)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Straubeberg, Ph.D.
Email: rgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: NDAM338 row: g column: 08
High quality sequence stop: 636.
Location/Qualifiers
1..843

```

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 843;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
40 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 92

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293239"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

Query Match 6.9%; Score 46.6; DB 6; Length 843;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
40 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 92

RESULT 11
CB588525 859 bp mRNA linear EST 03-APR-2003
LOCUS AGENCOURT_12567944 NIH_MGC_136 Mus musculus cDNA clone
DEFINITION IMAGE:30289751 5', mRNA sequence.
ACCESSION CB588525
VERSION CB588525.1 GI:29506381
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 859)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM329 row: e column: 24
High quality sequence stop: 763.

FEATURES
Source
1. 859
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30289751"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

Query Match 6.9%; Score 46.6; DB 6; Length 843;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
58 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 110

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 859;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
59 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 111

Corp. Note: this is a NIH_MGC Library."

Query Match 6.9%; Score 46.6; DB 6; Length 859;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
59 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 111

RESULT 12
CB589177 865 bp mRNA linear EST 03-APR-2003
LOCUS AGENCOURT_12770446 NIH_MGC_136 Mus musculus cDNA clone
DEFINITION IMAGE:30290534 5', mRNA sequence.
ACCESSION CB589177
VERSION CB589177.1 GI:29507033
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 865)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM331 row: f column: 15
High quality sequence stop: 630.

FEATURES
Source
1. 865
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30290534"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally, oligo-dT primed (5'-GACTAGTTCTAGATCGGAGCGGCCGCC(T)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by Resgen, Invitrogen Corp. Note: this is a NIH_MGC Library."

Query Match 6.9%; Score 46.6; DB 6; Length 865;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
58 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 110

Query Match 6.9%; Score 46.6; DB 6; Length 865;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 53
58 ATGGGAGCCTGGATTGTTGCTGCTGCTGAGGAGCAGCTTATGAC 110

IMAGE:30295123 5', mRNA sequence.
ACCESSION
CB587051
VERSION
CB587051.1 GI:29504907
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 886)
REFERENCE
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM343 row: e column: 20
High quality sequence stop: 619.
Location/Qualifiers
1. .886
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30295123"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPO6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTACTCTAGATCGGAGCGGCCGCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 886;
Best Local Similarity 92.5%; Pred. No. 0.85;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACAGCTTCACTATGCC 53
|||||
DB 84 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACAGCTTCTCTATGCC 136
|||||

RESULT 14
CB589251 889 bp mRNA linear EST 03-APR-2003
LOCUS
AGENCOURT_12780700 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30287069 5', mRNA sequence.
ACCESSION
CB589251
VERSION
CB589251.1 GI:29507107
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 889)
REFERENCE
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM342 row: f column: 20
High quality sequence start: 19
High quality sequence stop: 658.
Location/Qualifiers
1. .891
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30294719"
/tissue_type="embryonic limb, maxilla and mandible"

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM322 row: f column: 06
High quality sequence start: 23
High quality sequence stop: 581.
Location/Qualifiers
1. .889
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30287069"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_136"
/note="Vector: PCMV-SPO6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTACTCTAGATCGGAGCGGCCGCTT)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.9%; Score 46.6; DB 6; Length 889;
Best Local Similarity 92.5%; Pred. No. 0.86;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACAGCTTCACTATGCC 53
|||||
DB 78 ATGGGAGCCTGATTTGTTGCTGCTCTCGGAGACAGCTTCTCTATGCC 130
|||||

RESULT 15
CB574837 891 bp mRNA linear EST 02-APR-2003
LOCUS
AGENCOURT_12972279 NIH_MGC_136 Mus musculus CDNA clone
IMAGE:30294719 5', mRNA sequence.
ACCESSION
CB574837
VERSION
CB574837.1 GI:29494367
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 891)
REFERENCE
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDAM342 row: d column: 24
High quality sequence start: 19
High quality sequence stop: 658.
Location/Qualifiers
1. .891
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30294719"
/tissue_type="embryonic limb, maxilla and mandible"

/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5' 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dt primed (5'-GACTAGTCTAGATCGGAGCGGCCGCC(7)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 6.3%; Score 46.6; DB 6; Length 891;
Best Local Similarity 92.5%; Pred. No. 0.86;
Matches 49; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGGGAAGCTGATTTTGTTCCTGCTCTGGAGACAGCTTCAGTATGCC 53
|||
Db 79 ATGGGGAAGCTGATTTTGTTCCTGCTCTGGAGACAGCTTCAGTATGCC 131

Search completed: February 8, 2005, 22:18:22
Job time : 8100.44 secs